A MOLECULAR GENETIC INVESTIGATION INTO 2-METHYl-3-AMYL-6-METHOXYPRODIGIOSENE (PRODIGIOSIN) BIOSYNTHESIS IN *Serratia marcescens*

by

Sophia Y. Abbasi Bsc (Hons) University of Aston

A thesis presented for the degree of

Doctor of Philosophy

Department of Biological Sciences

University of Warwick

September 1998

CONTENTS

SUMMARY

Serratia marcescens is a Gram-negative enteric bacterium. The distinguishing feature of this species is the production of a bright red, non-diffusible pigment: 2-methyl-3-amyl-6-methoxyprodigiosene (prodigiosin). Prodigiosin is a classical secondary metabolite, produced in late-log to stationary phases of growth. It belongs to a family of structurally related tripyrrolic compounds, produced by a number of prokaryotic genera, which possess anti-bacterial, anti-fungal, anti-protozoal and immunosuppressive properties. The biological function(s) of prodigiosin in *Serratia marcescens* is unknown. Prodigiosin is synthesised from proline, alanine, serine, methionine, glycine and acetate. Little is know about the biosynthetic pathway, except that it is bifurcated, and terminated in the condensation of a bipyrrole and a monopyrrole to form prodigiosin. Virtually no published information exists on the pathway precursors, the pathway enzymes or the genes encoding them. The aim of this study was to investigate prodigiosin biosynthesis at the genetic level. Thomson (1996), isolated the prodigiosin biosynthetic (*pig*) gene cluster from a *S. marcescens* chromosomal DNA library. After subcloning, approximately half of the *pig* cluster (11.5 Kb) was sequenced in this study. The remainder was sequenced in a parallel study in this laboratory. From both studies, 16 putative open reading frames (ORFs) were identified, which are arranged unidirectionally, with the exception of *orf16* at the extreme 3' end of the cluster. In the 5' half of the cluster, sequenced in this study, homologues of bacterial acyl-CoA dehydrogenase, phosphoenolpyruvate synthase and ornithine aminotransferase were identified by similarity. A homologue of a hypothetical protein mapping to the *red* (undecylprodigiosn) locus of *Streptomyces coelicolor* A3(2) was also identified by similarity; another putative ORF does not have any database homologues. The pig cluster was randomly mutagenised by using Tn*phoA*'-2, which simultaneously generated some *lacZ* gene fusions strains. Non-pigmented, hyper-pigmented and orange-pigmented mutants were isolated. Cloning and sequencing of transposon insertions from non-pigmented Tn*phoA*'-2 mutated strains revealed that gene fusions to the first and fourth putative ORFs had been obtained. Additionally, it was found that in one non-pigmented mutant strain, a transposon insertion is located in an ORF encoding a putative homologue of the *Escherichia coli* integral inner-membrane histidine sensor kinase EnvZ; sequence data suggest that a putative homologue of the corresponding response-regulator OmpR is present upstream of the transposon insertion site. Another non-pigmented mutant was found to have a transposon insertion in a putative homologue of *Escherichia coli hscA*, which encodes a "cold-shock" induced molecular chaperonin. Southern blot analyses showed that insertions in these latter coding regions are external to the pig cluster. Two hitherto unknown loci, which are essential to pigment biosynthesis, were therefore identified in this study. Strains carrying gene fusions in *orf1* and *orf4* of the pig cluster showed differential LacZ expression under prodiogiosin biosynthesis-permissive conditions. LacZ expression was not abolished by growing these strains at a temperature at which prodigiosin biosynthesis does not occur, suggesting that transcription of *orf1* and *orf4* is not temperature-sensitive. Other work done in this study included strain construction by the use of *a Serratia marcescens* generalised transducing phage (ϕOT8), and the construction of a Lac^- strain of this species in which lacZ fusions were generated by Tn*phoA*'-2 mutagenesis, and transduced into from an isogenic Lac^+ strain.

LIST OF FIGURES

CHAPTER 7

LIST OF TABLES

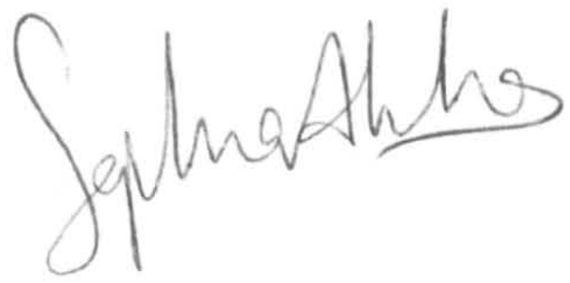
ACKNOWLEDGEMENTS

I would like to thank my supervisor George Salmond for giving me the opportunity to work in bacterial molecular genetics, despite the fact that I was an uninitiated human biologist. I also thank "my post-doc" Tony Cox for all the groundwork which allowed this project to happen and his help over the past few years. A big thank-you goes out to all the Salmond group, past and present who helped me learn-on-the-job: Nick Thomson (especially you), Steve Harris, Susie Wharam, Simon McGowan, Oona Harrison, Paul Wan and lateral-thinker extraordinaire Ste Bentley. Thanks to my best mate Manyee for everything and my other girly-night-out accomplices Michèle and Kathryn for good times. No thanks at all to Neil Whitehead and Dan Todd for their pearls of wisdom on every aspect of life, straight from the pages of "The Currant Bun". Thank-you to Christine and Lubna for being my writing-up buddies. A heartfelt thank-you to my family, especially my mum, and other friends for all their support and encouragement. Finally, thank-you to David for constant love and support.

I acknowledge financial support for this work from the BBSRC.

DECLARATION

This thesis has been compiled by myself and has not been used in any previous application for a degree. The results were obtained by myself, except where the contributions of others has been acknowledged. All sources of information have been specifically acknowledged by means of references.



Sophia Yasmin Abbasi

September 1998

ABBREVIATIONS

Φ	bacteriophage
λ	bacteriophage lambda
ΔG	change in free energy
μg	micrommes
μl	microlitres
μM	micromolar
A	absorbance
A	adenine
Å	angstrom(s)
A.	*Archaeoglobus*
AMPS	ammonium persulphate
Ap	ampicillin
ATP	adenosine triphosphate
B.	*Bacillus*
BlaM/ *blaM*	mature β-lactamase
bp	base pairs
Bre.	*Brevibacillus*
BSA	bovine serum albumin
C	cytosine
°C	degrees centigrade
C-terminal	carboxy-terminal
CAP	calf intestinal alkaline phosphatase
Car/*car*	1-carbapen-2-em-3-carboxylic acid (carbapenem)
cm	centimetre(s)
Cm	chloramphenicol
CoA	coenzyme A
conc	concentrated
Da	Daltons
dH_2O	distilled water

DIG	digoxigenin
DMF	dimethylformamide
DNA	deoxyribonucleic acid
ds	double-stranded
DTT	dithiothreitol
dATP	2'deoxyadenosine 5'-triphosphate
dCTP	2'deoxycytidine 5'-triphosphate
dGTP	2'deoxyguanosine 5'-triphosphate
dNTPs	2'deoxynucleotide 5'triphosphates
dTTP	2'2'deoxythymidine 5'-triphosphate
E.	*Escherichia*
EDTA	diaminoethanetetra-acetic acid
EMS	ethylmethane sulphonate
Ent.	*Enterococcus*
Erw.	*Erwinia*
EtBr	ethidium bromide
EtOH	ethanol
FAD	flavine adenine dinucleotide
Fig.	figure
g	gramme(s)
h	hour(s)
HIV	human immuno-deficiency virus
HSL	homoserine lactone
IPTG	isopropyl β-D-thiogalactopyranoside
IS	insertion sequence
Kb	kilobase pair(s)
KDa	kilodalton(s)
Kn	kanamycin
kV	kilovolts
L	litre(s)
LacZ/*lacZ*	β-galactosidase
LB	Luria Bertani medium

LBA	Luria-Bertani medium agar
LTTR	LysR type transcriptional regulator
M	moles or molar
M.	*Megasphaera*
MAP	2-methyl-3-amylpyrrole
mA	milliampere(s)
MBC	4-methoxy-2,2'-bipyrrole-5-carboxyaldehyde
Met.	*Methanococcus*
mg	milligramme(s)
min	minute(s)
ml	millilitre(s)
mm	millimetre(s)
mM	millimolar
MM	minimal medium
MMA	minimal medium agar
moi	multiplicity of infection
mRNA	messenger RNA
Mta.	*Methanobacteruim*
MW	molecular weight
Myx.	*Myxococcus*
N-terminal	amino-terminal
NADH	nictotinamide adenine dinucleotide, (reduced form)
NaOAc	sodium acetate
NB	nutrient broth
NBA	nutrient broth agar
nt	nucleotides
OAT	ornithine aminotransferase
OD	optical density
OHHL	*N*-(3-oxohexanoyl-L-homoserine lactone)
ORF/Orf	Open reading frame
PAGE	polyacrylamide gel electrophoresis
PEP	phosphoenol pyruvate

Pers. comm.	personal communication
P_i	inorganic phosphate
pI	isoelectric point
Pig/*pig*	pigment [prodigiosin]
PhoA/*phoA*	alkaline phosphatase
Pse.	*Pseudomonas*
psi	pounds per square inch
RB	running buffer
RBS	ribosome binding site (equivalent to the Shine-Dalgarno sequence only)
Rf	rifampicin
S.	*Serratia*
sdH_2O	sterile distilled water
S-D	Shine-Dalgarno
SDS	sodium doecyl sulphate
Sp	spectinomycin
Stp.	*Staphylococcus*
SSC	standard sodium citrate
Sta.	*Staphylothermus*
Str.	*Streptomyces*
ss	single-stranded
subsp.	subspecies
t	time
T	thymine
TBE	Tris-borate-EDTA buffer
Tc	tetracycline
TE	Tris-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
Tn	transposon
Tris	2-amino-2(hydroxymethyl)1,3-propane diol
U	uracil

CHAPTER 1

INTRODUCTION

1.1.1 HISTORICAL ASPECTS OF *Serratia marcescens* MICROBIOLOGY

The genus name *Serratia* is one of the oldest in bacterial nomenclature. It was given by scientist Bartolomew Bizio, who in 1823 reported that a minute stemless fungi, clustered in semispherical capsules, had caused reddening or bleeding of warm moist polenta. The name was in tribute to Serafino Serrati, an Italian physicist who, according to Bizio, invented and sailed the world's first steamboat on Florence's River Arno (Gaughran, 1969). Although not then identified as a bacterium, the name was subsequently retained for the bacterial genus. The present species name *marcescens*, meaning "fading away", is derived from the Latin verb *marcesco*, and pertains particularly to decay or putrefaction (Williams and Qadri, 1980). Bizio's report of bleeding polenta was later credited as the first reported description of a specific change effected by a microbe (Thimann, 1963).

1.1.2 THE GENUS *Serratia*: TAXONOMY, DISTRIBUTION AND CHARACTERISTICS

Serratia belongs to the Enterobacteriacaea family and is therefore closely related to *Escherichia coli*. Members of the genus are non-spore-forming rods, 0.9-2μm in length and 0.5-0.8μm in diameter, and they are motile by means of peritrichous flagella. *Serratia* spp. are facultatively anaerobic (Grimont and Grimont, 1984). The mol % G + C content of *Serratia* is 52-60%, with a mean value 58.4%; S. *marcescens* has the highest G + C content among enteric bacteria (Grimont and Grimont, 1984). The genome size of *Serratia marcescens*, the type species of the genus, based on analysis of one strain, is estimated to be 3.57×10^9, which is 10^9 Da larger than that of *E. coli* (Grimont and Grimont, 1978). Certain species in this genus are able to produce the characteristic, prodigiosin (section 1.3).

Previously, many more species were included in the genus *Serratia* (Grimont *et al*., 1977; Grimont and Grimont 1978), but the number of species assigned to the genus currently numbers ten, comprising the following: *S. entomophila*, *S. ficaria*, *S. fonticola*, *S. grimesii*, *S. liquefaciens*, *S. marcescens*,

S. odifera, *S. plymuthica*, *S. proteomaculans*, and *S. rubidaea* (Grimont and Grimont, 1984). Only *S. marcescens*, *S. rubidaea and S. plymuthica* synthesise prodigiosin.

Serratia spp. grow well in the 20-37°C temperature range; *S. liquefaciens*, *S. odifera* and *S. rubidaea*, grow readily at 4-5°C (Grimont and Grimont, 1984). Members of the genus are distributed ubiquitously in air, soil, freshwater and sea water. Habitats from which they have been isolated include grass, mushrooms, and commercial vegetables and salad crops as well as man-made materials (Grimont and Grimont, 1978).

Not surprisingly, *Serratia* spp. produce an battery of exoenzymes which allow colonisation of diverse habitats. These include serine protease, metalloprotease, nucleases, lipases, chitinase and cellulase (Yanagida, 1986; Braun and Schmidtz, 1980; Clegg and Allen, 1985; Hines, 1988; Grimont and Grimont, 1984). All species are capable of hydrolysing DNA, gelatine, and solubilised casein. Chitin is hydrolysed by all strains except *S. odifera* and *S. rubidaea* (Grimont and Grimont, 1984).

Serratia spp. have the ability to colonise surfaces by three means. Firstly, "swimmer" cells are typically short rods, 0.9-2μm in length with one or two flagella. In environments of high microviscosity, swimmer cells have the ability to undergo morphogenesis by cellular elongation to 5-30μm and become hyperflagellated, converting into swarmer cells (Alberti and Harshey, 1990). Some *Serratia* spp., including *S. marcescens*, are also capable of producing a surface-active wetting agent called Serrawetin (cyclo-[D-3-hydroxydodecanoyl-L-seryl]$_2$) which enables swarming cells to undertake spreading growth. This may be a virulence determinant (Matsuyama *et al.*, 1995).

1.1.3 PATHOGENICITY

The ability of *Serratia* spp. to cause disease varies greatly not only between species, but also between biotypes. In animals other than humans, *Serratia* spp. can be an aggressive pathogens. *Serratia* spp. are insect pathogens, infecting a range of genera, although this is only considered to be a serious threat for reared insects (Grimont and Grimont, 1978). *Serratia* spp. are also associated

animals. They have been isolated from cold-blooded vertebrates (Grimont and Grimont, 1978). Poultry birds have been found to carry *Serratia* spp. in their digestive tracts, occasionally resulting in infection of eggs, death of embryos and carcass contamination (Grimont and Grimont, 1978). *Serratia marcescens* is a well-known pathogen of cattle in which it causes mastitis (Grimont and Grimont, 1984; Kamarudin *et al*., 1996), and therefore it is commercially important in this respect. *Serratia*-induced fatal septacaemia has been reported in pigs and goats and horses (Grimont and Grimont, 1978). *Serratia* spp. have also been isolated from fish and shellfish (Grimont and Grimont, 1978).

In humans, *Serratia* spp. are generally an opportunistic pathogens, with *S. marcescens* being responsible for the great majority of infections (Grimont and Grimont, 1984; Hejazi and Falkiner, 1997). Advances in transplant medicine and cancer-therapy and also the emergence of HIV, have resulted in increasing numbers of immunosuppressed patients and this has also been paralleled by increased incidence of *S. marcescens* infections in recent decades; infections are usually acquired nosocomially (Grimont and Grimont, 1978; Hejazi and Falkiner, 1997). Earlier in the century, the species was considered avirulent in humans. In fact, it was used as a trace organism in clinical experiments due to production of prodigiosin (Simberkoff, 1980). One of the main reasons why *S. marcescens* is able to flourish in an opportunistic manner is due to its ability to colonise surfaces such as medical equipment and devices including prostheses and catheters (Daschner, 1980; Hejazi and Falknier, 1997) and even contact lenses (Parment, 1997). *S. marcescens* can cause serious infections in neonates including meningitis, septicaemia and conjunctivitis (van Ogtrop *et al*., 1997).

Intriguingly, clinical isolates of *S. marcescens* are usually non-pigmented whereas colonisation of body tissues by pigmented strains is usually asymptomatic (Williams and Qadri, 1980; Ding and Williams, 1983). Pigmented strains are responsible for "red diaper syndrome" in infants which has been reported in medical literature (Waisman, 1958, cited in Gaughran, 1969; Grimont and Grimont, 1978). This syndrome is due to chronic intestinal *S. marcescens* infection, and manifests as the development of blood-red coloration on soiled nappies when they were stored prior to washing.

The reason for the difference in pathogenicity between pigmented and non-pigmented strains of *S. marcescens* has not been established. Various studies have shown that that no apparent selective advantage is conferred on non-pigmented cells of *S. marcescens* in comparison to pigmented cells, with both having the same growth kinetics, total protein yield and oxygen uptake when cultured *in vitro* (Williams *et al.*, 1971b). It has been suggested that non-pigmented clinical isolates are not cultured for sufficiently long for pigment to appear (Grimont and Grimont, 1978). However, Ding and Williams (1983), showed by cross-feeding studies that such clinical isolates are rarely able to produce prodigiosin, being defective in stages of the biosynthetic pathway.

1.2.1 PIGMENTS IN BACTERIA

Prokaryotes synthesise a diverse array of pigments, with many different functions, which play numerous roles in their respective hosts (reviewed in Feistner, 1990). Well characterised pigments include carotenoids, produced by, for example, *Myxococcus xanthus*,which protect the cell from light-induced damage (Armstrong, 1994). Other pigments, such as bacteriochlorophyll are used in photosynthesis by cyanobacterial species. Pigments can also be antibiotic, for example actinorhodin and undecylprodigiosin (also called Red), which are blue and red respectively, produced by *Streptomyces coelicolor* A3(2).

1.2.2 THE PRODIGIOSINS

The prodigiosins are tripyrrolic pigment compounds which invariably contain an aromatic portion with an *O*-methyl side chain. Several nomenclatures have been suggested in the literature for classifying these pigments (reviewed in Williams and Qadri, 1980). Classification is based on the linear tripyrrole prodigiosin as being the archetype pigment in this family. Related compounds, produced by other prokaryotic genera, are usually referred to in the literature as prodiginines (Gerber and Lechevalier, 1976) and are characterised by having complex side chains and modifications. The pigments are also collectively (and informally) termed "prodigiosins". Identification of prodigiosin and similar pigments exploits their characteristic absorption spectra in acidified ethanol, as

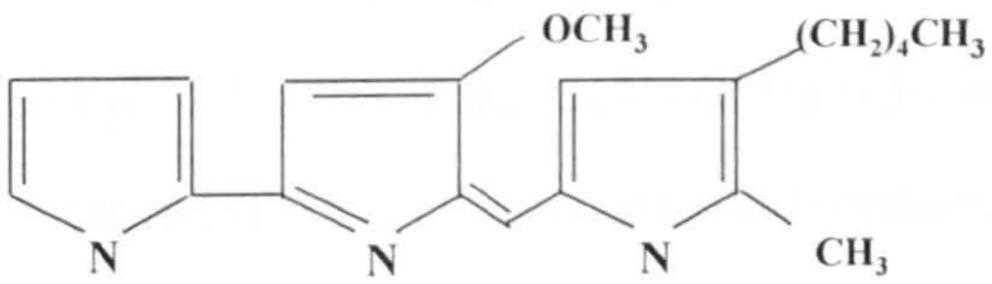
they possess absorption maxima between 510 to 570 nm and a characteristic shoulder ~35 nm lower (Williams and Qadri, 1980). Outside the genus *Serratia*, prodigiosins have been isolated from a number of actinomycetes (most notably *Str. coelicolor* A3(2)), *Pseudomonas magnesorubra* and *Vibrio psychoerythrus* (Gerber and Lechavalier, 1976). The structures of prodigiosin and some prodiginines are shown in Fig. 1.1.

Prodigiosins have been the subject of many decades of research. They have antiprotozoal, antifungal and antibacterial properties (Gerber, 1975; Grimont and Grimont, 1984). They are also immunosuppressive, being active against cultured mammalian cell lines, e.g. T-lypmhocytes, and also in animal models (Lee *et al*., 1997; Magae *et al*., 1996; Tsuji *et al*., 1992; Katoaka *et al*., 1992). Prodigiosin was also produced commercially at one time (Nager, 1951, cited in Gerber, 1975).

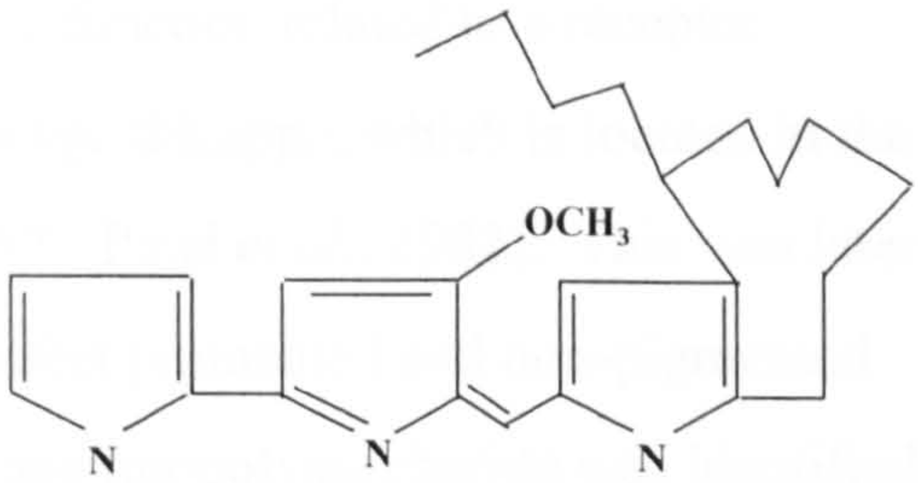
1.3.1 PRODIGIOSIN AND *Serratia marcescens*

Prodigiosin biosynthesis has been most extensively studied *in S. marcescens*, and production in other *Serratia* spp. shall not be considered here. The production of this highly characteristic blood-red pigment is the single most conspicuous characteristic which has allowed the history of *S. marcescens* to be traced back to ancient times:. the compound's name is inextricably linked with the bacterium's history, and is believed to have been coined by Kraft in 1902 (cited in Williams and Qadri, 1980). It is derived from the former species synonym *Bacillus prodigiosus*. The precise derivation of the suffix *prodigiosus* cannot be traced for certain, but is thought to have been applied because of the bacterium's possible associations with numerous historical, religious, and seemingly miraculous happenings throughout the history of Western Civilisation (reviewed in Gaughran, 1969, Cullen, 1994 and Thomson, 1996). The trivial name is used in this study, and the formal name for the compound is, according to the nomenclature of Hearn *et al*.(1970), 2-methyl, 3-amyl, 6-methoxy-prodigiosene.

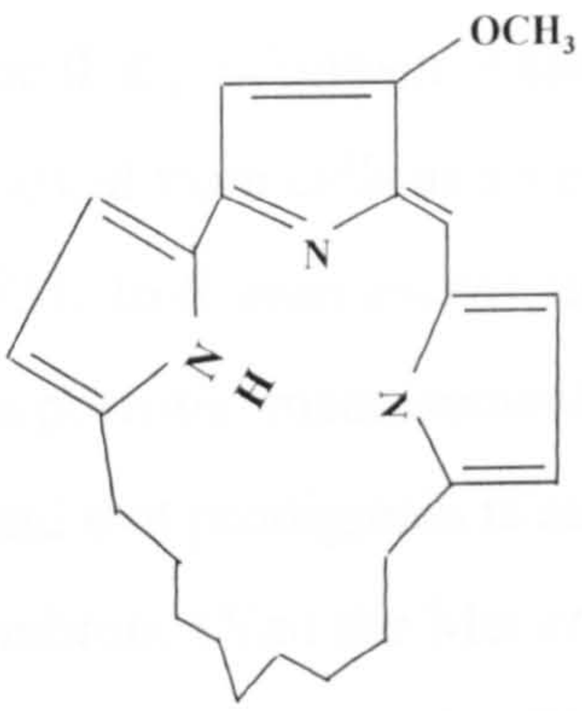
FIGURE 1.1



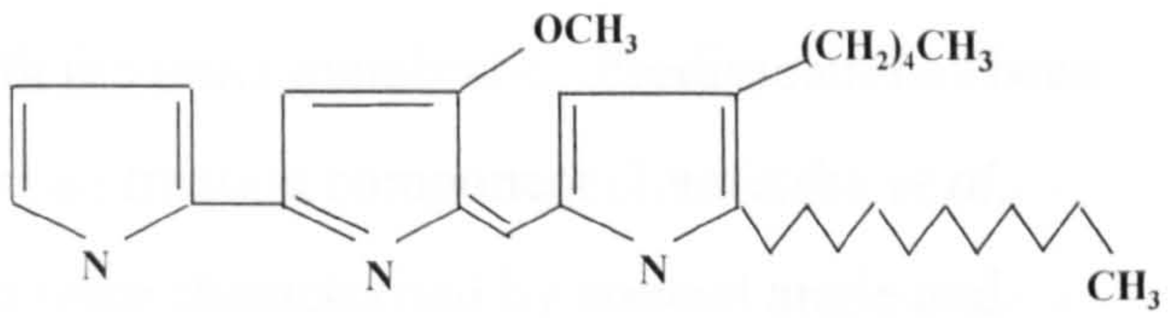
(I)



(II)



(III)



(IV)

FIGURE 1.1 THE STRUCTURE OF PRODIGIOSIN AND SOME RELATED PRODIGININES

The pigment are tripyrroles. (I) prodigiosin produced by *S. marcescens*; (II) butylcycloheptylprodiginine, (III) cyclononylprodiginine, (IV) undecylprodigiosin. Compounds II-IV are produced by actinomycetes. Undecylprodigiosin (IV) is the major constituent prodiginine in Red, produced by *Streptomyces coelicolor* A3(2) and compound III is the minor constituent. (After Williams and Qadri, 1980 and Gerber and Lechevalier, 1976).

Studies have shown that prodigiosin is a non-diffusible membrane-bound pigment in *S. marcescens*. As early as 1969, Hubert *et al*. stated that the cell envelope seems to be necessary as a physical site for prodigiosin biosynthesis. Its precise location in the cell envelope however, has been a subject of some debate. It was originally proposed that prodigiosin is associated with the outer membrane, because it was found that SDS treatment of pigmented cells liberates prodigiosin together with outer membrane components (Tsang *et al*., 1971). It was also proposed that prodigiosin may have a function related to a receptor complex for the *S. marcescens* transducing phage ΦKappa, which is located in the outer membrane (Paruchuri and Harshey, 1987; Patel *et al*., 1982). This was later disproved as it was found that ΦKappa can infect pigmented and non-pigmented strains with equal voracity, and outer membrane lipopolysaccharide was identified as the bacteriophage receptor (Montilla *et al*., 1991). Currently, the opinion seems to be that prodigiosin is associated with the inner membrane. Prodigiosin has been removed from cells as an inner membrane fraction component (Lauferska *et al*., 1983). In *S. marcescens* strains which were characterised by contact angle and zeta potential measurements, together with various forms of spectroscopy, it was found that prodigiosin is associated with a cellular layer deeper than the outer membrane (Van der Mei *et al*., 1991). Other studies have found that pigment is associated with protein (Woods *et al*, 1971). A more recent study has identified a 100 KDa protein associated with pigment which is not present in non-pigmented cells; this protein was found exclusively in the cytoplasmic/inner membrane fraction of disrupted cells (Kobayashi and Ichikawa, 1989)

Prodigiosin is classical secondary metabolite in *S. marcescens*, being produced in late-log to stationary phases of growth (Williams, 1973). Despite many years of research, the physiological function of the pigment in *S. marcescens* remains an enigma. Possible functions attributed to prodigiosin are discussed in the next section.

1.3.2 POSSIBLE ROLES OF PRODIGIOSIN IN *Serratia marcescens*

Many studies have attempted to define a role for prodigiosin in this species. Firstly, the antibiotic properties of the pigment, mentioned previously

(section 1.2.2) would confer obvious selective advantages. It has also been observed that prodigiosin is a potent photosensitiser when released, which may be lethal to competitor species (Williams and Qadri, 1980). It is suggested that the hydrophobic nature of the pigment facilitates dispersal and adherence to various surfaces, including plastics which could conceivably aid niche colonisation and therefore confer a selective advantage (Burger, 1985). However, it was found in another study that hydrophobicity is also due to other unidentified factors common to both pigmented and non-pigmented variants of the species (Rosenberg, 1984). In a study of physiochemical cell surface properties of *S. marcescens*, it was found that surface hydrophobicity does not differ between pigmented and non-pigmented strains (Van der Mei *et al.*, 1992). Colour variation is reportedly associated with a change in surface antigens, which has been suggested to help evade a host immune response in pathogenesis (Paruchuri and Harshey, 1987). This however conflicts with the observation that most clinical isolates of *S. marcescens* are non-pigmented. Another hypothesis for the role of prodigiosin is that its biosynthesis provides a means for discarding excess reducing equivalents in senescent cultures, under conditions unfavourable for growth, thereby supporting the oxidative activity of the cells at the requisite level. In a study to test this, the addition of chemical shunting agents, e.g. phenazine methosulphate, was accompanied by a decrease in prodigiosin biosynthesis (Trutko and Akimenko, 1989). However, as mentioned previously, no differences in cell viability or other parameters, between pigmented and non-pigmented strains have been observed when cultured under the same conditions (Williams and Qadri, 1980). In the analagous system of Red (undecylprodigiosin) biosynthesis in *Str. coelicolor* A3(2), Red is believed to act as a shunt for excess intracellular proline (Hood *et al.*, 1992). This is discussed in greater detail in section 1.3.5.2.

1.3.3 CURRENT KNOWLEDGE OF THE PRODIGIOSIN BIOSYNTHETIC PATHWAY IN *Serratia marcescens*

Although prodigiosin has been the subject of numerous studies through the century, relatively little is known the biosynthetic pathway to its production. Particularly pertinent to this study, virtually no published information exists on the

genes encoding the pathway enzymes. Work done in this laboratory has shed some light on regulation of prodigiosin production at the genetic level (discussed in section 1.3.4.2).

Studies have shown that the pyrrole rings of prodigiosin arise from amino acids and acetate rather than through the succinate-glycine cycle, as is characteristic of the pyrrole groups of porphyrins (Williams, 1973). Nuclear Magnetic Resonance (^{13}C) experiments established that prodigiosin is synthesised from acetate, glycine, serine, alanine and methionine, and proline, with the latter molecule being incorporated intact into prodigiosin (Wasserman *et al*., 1973, Fig. 1.2). Biochemical and cross-feeding studies using mutants determined to be blocked in different stages of the pathway revealed that prodigiosin is synthesised via a bifurcated pathway (Morrison, 1966; Williams and Qadri, 1980). The pathway is summarised in Fig 1.3. One branch of the pathway terminates in synthesis of the bipyrrole 4, methoxy-2,2'-bipyrrole-5-carboxyaldehyde (MBC), and the other branch terminates in synthesis of the monopyrrole, 2-methyl-3-amylpyrrole (MAP); these components are enzymatically condensed to form prodigiosin. (Morrison, 1966). The only intermediate to MBC or MAP which has been identified is 4, hydroxy-2,2'-bipyrrole-5-carboxyaldehyde (HBC, Fig. 1.3). It was found that some biosynthetic mutants did not synthesise prodigiosin (unless furnished with required intermediates), but were still pigmented. Mutant OF (Fig. 1.3) produces the orange pigment norprodigiosin, resulting from the condensation of HBC and MAP. Mutants producing a deep blue or intensely purple pigment, have also been previously isolated; this pigment believed to be dipyrroledipyrromethane and results from condensation of two MBC moieties (Gerber, 1975; Williams and Qadri, 1980).

Prior to work done in this laboratory, the last published attempt to clone the genes encoding prodigiosin biosynthesis was made in 1984 by Dauenhauer *et al*.. In that study, *E. coli* was transduced using a *S. marcescens* cosmid library. Pigment production was not reconstituted in *E. coli* but the work demonstrated that the prodigiosin biosynthetic genes are located on the *S. marcescens* chromosome and by supplying either or both of MAP and MBC, some

Figure 1.2

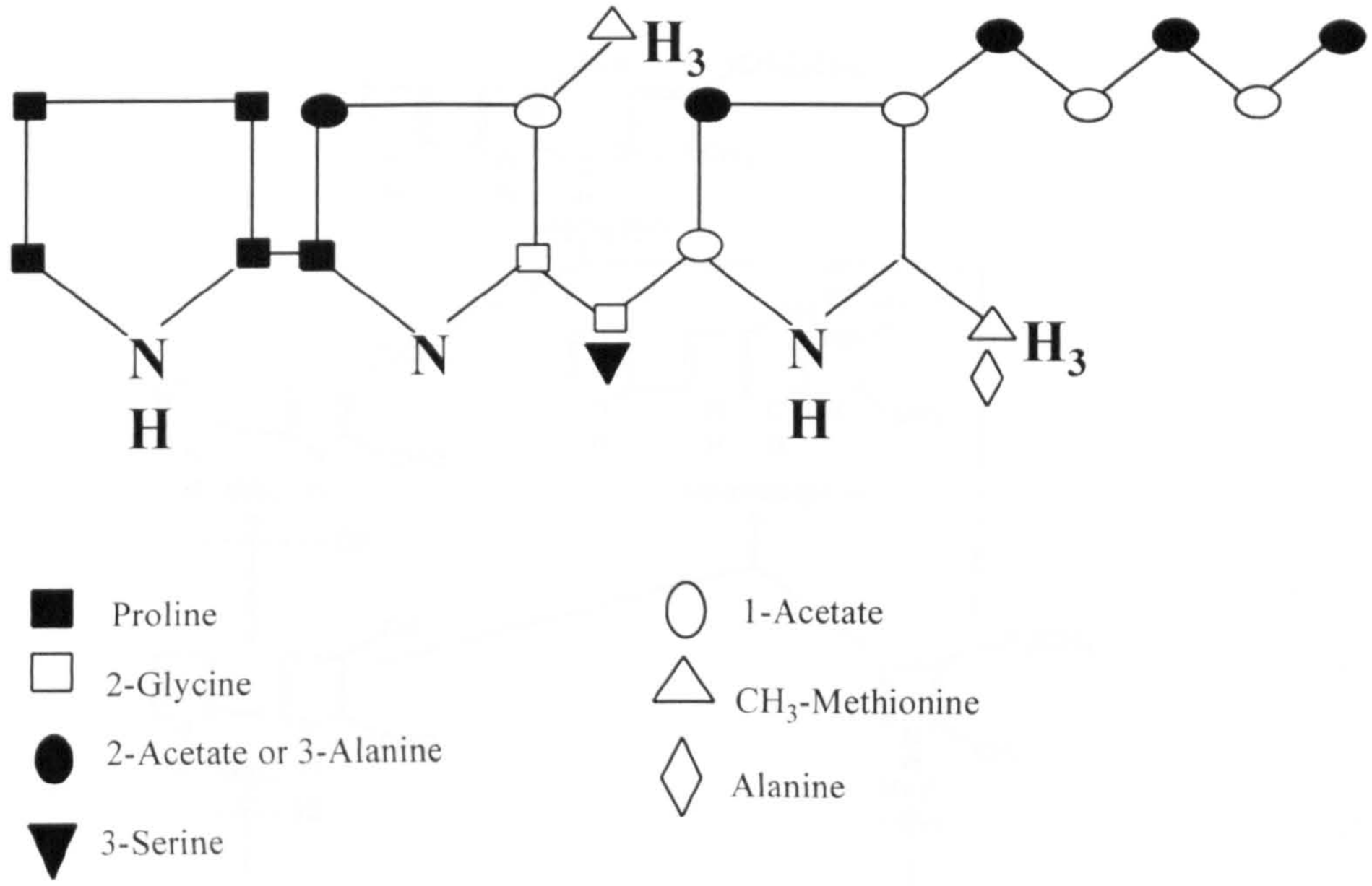


FIGURE 1.2 PATTERNS OF INCORPORATION OF AMINO ACID PRECURSORS INTO THE PRODIGIOSIN MOLECULE IN *Serratia marcescens*

Numbers or group preceding the ^{13}C labelled substrates indicate the carbon atom in that molecule which was found to be incorporated into prodigiosin. Taken from Williams and Qadri, 1980, based on studies by Wasserman *et al*., (1973)

Figure 1.3

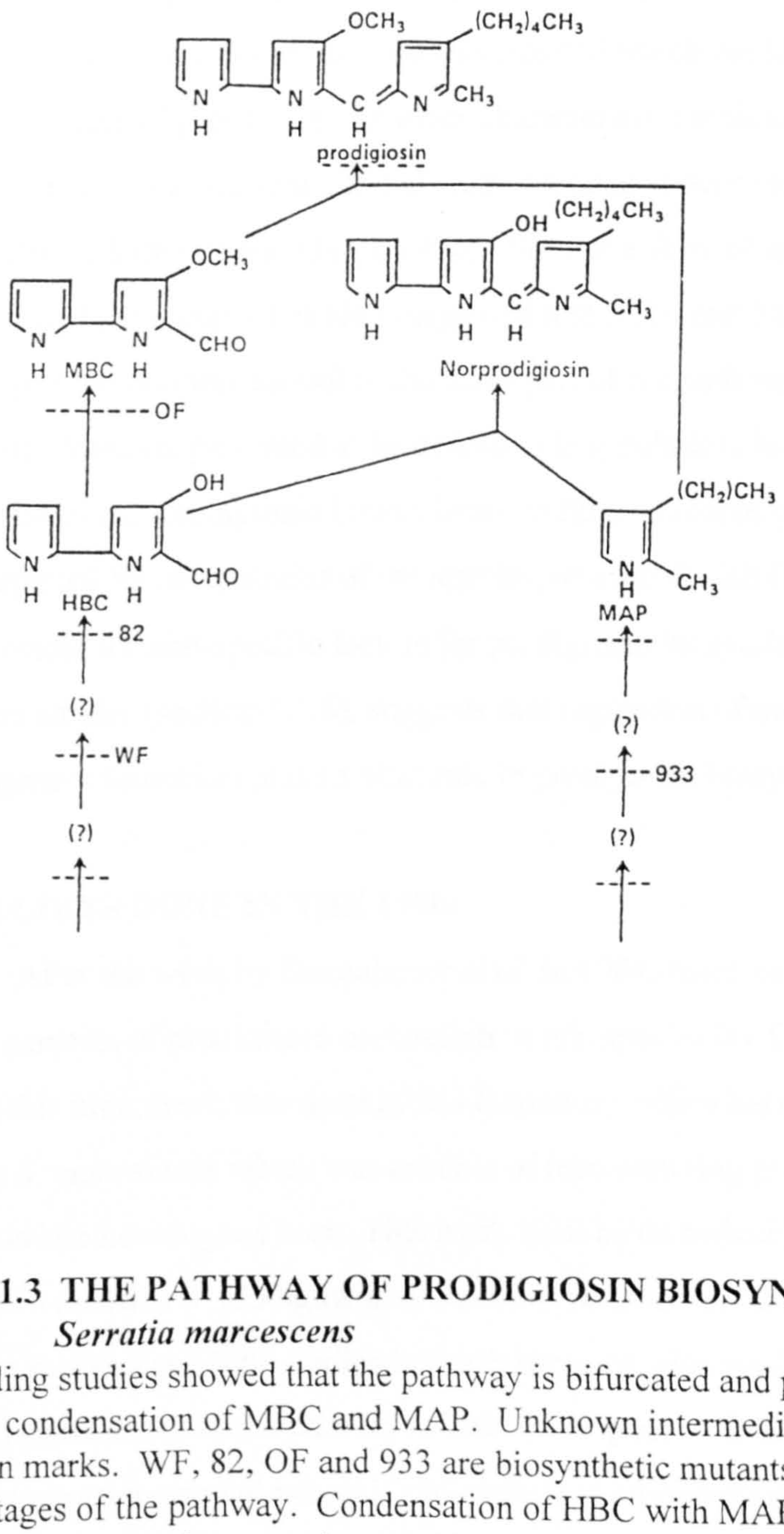


FIGURE 1.3 THE PATHWAY OF PRODIGIOSIN BIOSYNTHESIS IN *Serratia marcescens*

Cross-feeding studies showed that the pathway is bifurcated and prodigiosin is formed by condensation of MBC and MAP. Unknown intermediates are denoted by question marks. WF, 82, OF and 933 are biosynthetic mutants blocked at different stages of the pathway. Condensation of HBC with MAP in mutant OF resulted in formation of orange pigment (2-methyl-3-amyl hydroxyprodigiosin, norprodigiosin). Taken from Williams and Qadri, 1980.

of the cosmids could direct prodigiosin biosynthesis in the heterologous host system.

1.3.4 CURRENT KNOWLEDGE OF THE REGULATION OF THE PRODIGIOSIN BIOSYNTHETIC PATHWAY IN *Serratia marcescens*

1.3.4.1 STUDIES DONE PRIOR TO THE 1990s

Very little published information exists in this area. In syntrophic cross-feeding studies, a class of mutants was isolated which could not act as donors or acceptors of precursors, but other characteristics indicated that this class was blocked in a single step that exerted common control over synthesis of both MAP and MBC (Williams and Qadri, 1980). Similar effects of temperature, anaerobiosis and antimetabolites also suggested that MAP and MBC syntheses may be subject to common control in the early part of the pathway (Williams and Qadri, 1980). Mutants presumed to be defective in regulation, have been isolated and it was found that prodigiosin biosynthesis could be restored by unidentified factor(s) supplied by other strains of the species, or even *E. coli* (Rizki, 1954). The requirement for non-specific factors for prodigiosin biosynthesis, as inferred from certain studies (section 1.3.6), suggests that regulation of metabolism prior to onset of pigment formation plays a vital role in prodigiosin biosynthesis

1.3.4.2 STUDIES DONE IN THE 1990s

After the work by Dauenhauer *et al*. in 1984, research into the molecular genetics of prodigiosin production in this species lay fallow until the 1990s. At this time, work was done in this laboratory which led to the isolation of DNA from *S. marcescens* which was capable of reconstituting prodigiosin biosynthesis in a heterlogous host. This work shall be described here.

In addition to prodigiosin, *S. marcescens* strain ATCC39006, which was used in this and previous studies in this laboratory, also produces the β-lactam antibiotic 1-carbapen-2-em, 3-carboxylic acid (carbapenem, Car; Parker *et al*., 1982). In contrast to prodigiosin, this molecule is synthesised constitutively in ATCC39006 (Bycroft *et al*., 1987). Carbapenem is also produced by the closely

related Gram-negative enteric species *Erwinia carotovora* subsp. *carotovora*, which has also been the subject of investigation in this laboratory (McGowan *et al.*, 1995; Chhabra *et al.*, 1992; Parker *et al.*, 1982). Unlike in *S. marcescens*, Car production in *Erw. carotovora* subsp. *carotovora* is growth-phase dependent and under the control of a small diffusible global regulator molecule, *N*-3-oxohexanoyl-L-homoserine lactone (OHHL, Bainton *et al.*, 1992, see also section 1.4.3). To isolate and clone *S. marcescens car* genes, the cosmid library of ATCC39006 was transduced into *Erw. carotovora* subsp. *carotovora*. Transduction into *Erw. carotovora* subsp. *carotovora* mutants defective in *biosynthesis* of Car resulted in expression of *S. marcescens car* genes. In contrast, transduction into *Erw. carotovora* subsp. *carotovora* mutants defective in *positive regulation* of Car (i.e., mutants unable to synthesise OHHL, defective in the gene *carI*), did not result in expression of *S. marcescens car* genes in the same way (Cox, 1995). The *S. marcescens car* biosynthetic cluster was cloned and sequenced by Cox (1995; Cox *et al.*, 1998), who identified a *carR* homologue upstream of the *car* biosynthetic cluster. *carR* encodes the reciprocal response-regulator in the OHHL-dependent global regulation loop in *Erw. carotovora* subsp. *carotovora* (McGowan *et al.*, 1995). In *Erw. carotovora* subsp. *carotovora*, *carR* is autoregulated through CarR interaction with OHHL, and it is a cell-density dependent transcriptional activator which is part of the *luxR* family of transcriptional regulators (McGowan *et al.*, 1995). However, it was found that *S. marcescens carR* is OHHL-*in*dependent (Cox *et al.*, 1998), which is thought to explain why Car production is constitutive in this species.

Transduction of the *S. marcescens* cosmid library into certain strains of *Erw. carotovora* subsp. *carotovora* serendipitously resulted in the generation of blood-red pigmented colonies; the pigment was spectrophotometrically analysed and shown to be prodigiosin (Thomson, 1996). Thus, prodigiosin biosynthesis had been reconstituted, for the first time ever, in a heterologous host. Concurrently, Thomson (1996) mutagenised *S. marcescens* ATCC39006 using ethylmethane sulphonate, and isolated mutants which were pleiotropically defective in prodigiosin and carbapenem production. He subsequently showed that these mutants are defective in production of a single regulatory protein, Rap

(**R**egulation of **a**ntibiotic and **p**igment), which was located on a separate cosmid, pNRT300. Rap was found to have a homologue in *Erw. carotovora* subsp. *carotovora*, which was designated Hor (**H**omologue **o**f **R**ap). It was found that Rap and Hor are a novel subgroup in a growing superfamily of proteins regulating diverse physiological processes in pathogenic bacteria (Thomson *et al*., 1997). Interestingly, it was found that prodigiosin production in *Erw. carotovora* subsp. *carotovora* is only possible in an $OHHL^+$ (*carI*$^+$) background, despite the fact that there is no evidence for an OHHL-like molecule in *S. marcescens* strain ATCC39006 (Cox, 1995). Pigment production in *Erw. carotovora* subsp. *carotovora* was due to one *S. marcescens* chromosomal library-derived complementing cosmid, pNRT104 (Thomson, 1996), which carries an estimated 35 Kb of *S. marcescens* DNA, suggesting that the prodigiosin biosynthetic genes are clustered, which is a common phenomenon in secondary metabolite producers (Martin, 1992; see also sections 1.3.5.1 and 1.4.4). Attempts to reconstitute pigment production in *E. coli* using pNRT104 failed, even with the separate addition of *rap* (Thomson, 1996), which suggests that *Erw. carotovora* subsp. *carotovora* and *S. marcescens* contain additional, unidentified factor(s) which are required for pigment production. However, it is also possible that expression of *S. marcescens* genes in *E. coli* is poor. This has been reported previously by Yanagida *et al*., (1986), who found that the *S. marcescens* promoter for serine protease was inoperative in *E. coli*

Recent work by Harris *et al*. (1998) in this laboratory has identified HexA, another regulator of multiple virulence determinants in *Erw. carotovora* subsp. *carotovora*. This protein is homologous, in part, to the LysR-type transcriptional regulator (LTTR) family, which control diverse processes in bacteria (reviewed in Schell, 1993). Multicopy *hexA* of *Erw. carotovora* subsp. *carotovora* in was found to have global effects in *S. marcescens* when supplied *in trans*, resulting in decreased motility, exoenzyme production and also abolished prodigiosin production. Unlike most LTTRs, which are transcriptiptional activators (Schell, 1993), HexA exerts repressive effects (Harris *et al*., 1998). A cognate HexA species has not yet been identified in *S. marcescens*.

1.3.5.1 Red (UNDECYLPRODIGIOSIN) BIOSYNTHESIS IN *Streptomyces coelicolor* A3(2)

Comparatively, Red biosynthesis in *Streptomyces coelicolor* A3(2) has been investigated more extensively at the genetic level, in terms of biosynthesis and regulation. It is therefore necessary to briefly review the literature on this system, as these data were considered useful at the outset of this study. *Str. coelicolor* A3(2) is a Gram positive soil-dwelling bacterium, capable of producing exospores and exhibiting a complex life cycle. In addition to Red, this organism also produces an array of other secondary metabolites: actinorhodin (Act), methlenomycin (Mmy), a calcium-dependent antibiotic (Cda), a brown spore-associated pigment, and A-factor-like butyrolactone (see section 1.4.2), which is a regulatory molecule (Hopwood *et al*, 1995). The production of a red mycelial pigment in this species was first reported by Rudd and Hopwood in 1980. It was found to be a mixture of two prodiginines: butylcycloheptylprodiginine and undecylprodigiosin (Tsao *et al*., 1985; Fig. 1.1). This mixture is referred to as undecylprodigiosin, or Red (Takano *et al*., 1992).

Rudd and Hopwood (1980) isolated mutants defective in production of Red, which were grouped into five classes (RedA-E) on the basis of cosynthesis experiments, and studies showed that *S. coelicolor* Red mutants could cross-feed with *S. marcescens* prodigiosin mutants, indicating similar pigment biosynthetic pathways exist in the two species; it was deduced that RedA mutants corresponded to mutant WT of *S. marcescens* (Figure 1.1), and RedE mutants corresponded to mutant OF (Fig. 1.1; Feitelson and Hopwood, 1983). It was proposed that the RedE class of mutants (and therefore OF in *S. marcescens*), is defective in synthesis of an *O*-methyl transferase, which is required to catalyse the conversion of hydroxylated bipyrrole to methylated bipyrrole (Fig. 1.1) and it was later confirmed that *redE* (and additionally *redF*) is required for O-methyl-transferase synthesis (Feitelson and Hopwood, 1983). Mutants RedA-E were found to be defective in bipyrrole synthesis; Coco *et al*. (1991) isolated mutants defective in synthesis of the monopyrrole of undecylprodigiosin, which led to the isolation of mutants RedG -RedR. The MAP branch of the Red pathway is more complex than the monopyrrole pathway in *S. marcescens*, and this is believed to

be consistent with the production of two prodiginines in this species (Coco *et al.*, 1991). Red-encoding genes were eventually cloned out on a 35.7 Kb chromosomal fragment and it was predicted that the locus comprises at least eighteen *red* genes; this clustered gene organisation is typical of antibiotic biosynthetic genes in *Streptomyces*. (Malpartida *et al.*, 1990).

1.3.5.2 REGULATION OF Red BIOSYNTHESIS IN *Streptomyces coelicolor* A3(2)

It was found that the RedD class of mutants could not co-synthesise with any other class (Rudd and Hopwood, 1980; Feitelson *et al*, 1985), and *redD* cloned in multicopy resulted in overproduction of Red (Navra and Feitelson, 1990); RedD mutants were also shown to be incapable of expressing *redE* and *redF* (Feitelson *et al.*, 1985). These studies showed that RedD is a transcriptional activator of Red biosynthesis, and *redD* is located at one end of the *red* cluster (Navra and Feitelson, 1990). RedD is homologous to *actII*ORF4, a regulator of actinorhodin biosynthesis in the same species (Navra and Feitelson, 1990).

As mentioned previously, Red biosynthesis is believed to act as a shunt for excess intracellular proline. In bacteria, proline plays a role in osmoprotection. Osmoprotectants are low molecular weight solutes which the cell can accumulate without high toxicity, preventing dehydration in conditions of increased osmolarity. The major osmoprotectant in streptomycetes is proline (Killham and Firestone, 1984a, 1984b). It was found that mutants of *S. coelicolor* A3(2) which are defective in proline transport and catabolism, but not proline biosynthesis (Put), overproduced Red (Hood *et al.*, 1992). It is proposed that in this species, proline synthesis and degradation are in dynamic equilibrium; if the equilibrium is disturbed, Red biosynthesis serves to act as a sink for excess proline (Hood *et al.*, 1990).

RedD is a pathway-specific positive regulator. More recently RedZ, another regulator of Red biosynthesis, has been identified. RedZ is a protein which lies higher in the regulatory cascade, which is believed to influence Red biosynthesis by activating RedD; it was found that *redD* transcription can only occur in a $bldA^+$ background (White and Bibb, 1997). Bald (Bld) mutants of *Str.*

coelicolor A3(2) are those which are defective in production of antibiotics and aerial mycelium (Merrick, 1976). Sequencing revealed that certain Bld mutants contain a mutation in the gene *bldA*, which encodes the only leucine tRNA which can recognise the codon UUA (Lawlor *et al.*, 1987). The UUA codon is rare in this species, which has a high G + C content genome, and it was therefore proposed that the presence of this codon in certain genes acts a point of regulation, for example, the *actII*ORF4 gene whose product regulates Act biosynthesis, contains this codon, and Act biosynthesis is therefore *bldA*-dependent (Fernandez-Moreno *et al.*, 1991). Sequencing showed that *redD* does not contain a UUA codon. Subsequently, *redZ* was identified and sequenced. RedZ is homologous to the UhpA sub-family of transcriptional regulators and does indeed contain the rare UUA codon (Flaxman, 1995; Guthrie *et al.*, 1998). The transcriptional organisation of the right-hand end of the *red* cluster, as deduced from the studies discussed in this section, is shown in Fig. 1.5. Mutants have also been isolated that retain the ability to synthesise Red in a *bldA* background, provided they are cultured under low phosphate conditions. This phenotype is designated Pwb (**P**igmented **w**hilst **b**ald) (Guthrie and Chater, 1990). In a particular mutant, Pwb-6, it was found that a base change in the -35 promoter region of RedZ potentially causes a promoter-up mutation because the change makes the promoter sequence more similar to that of *E. coli* σ^{70} type promoter consensus (Flaxman, 1995; Guthrie *et al.*, 1998). This, together with low-level translation of the UUA codon by a charged non-cognate tRNA, is thought to explain why Pwb-6 is upregulated for Red biosynthesis.

1.3.5.3 CONTINUING STUDIES ON Red BIOSYNTHESIS IN *Streptomyces coelicolor* A3(2)

Genetic dissection of the *red* cluster should be facilitated by the systematic sequencing of the entire *Str. coelicolor* A3(2) genome, which is currently underway (Redenbach *et al.*, 1997). To date, this has resulted in the identification of other putative genes in the *red* cluster, and flanking chromosomal DNA. With the exception of *redD* and *redZ*, sequence data from the *red* cluster have, so far, only been published in database sources. These data shall be discussed as appropriate in subsequent chapters of this report.

1.3.6 FACTORS WHICH INFLUENCE PRODIGIOSIN BIOSYNTHESIS IN *Serratia marcescens*

S. marcescens is a non-fastidious and relatively fast-growing organism. Williams and Qadri (1980) measured the kinetics of prodigiosin biosynthesis, in relation to growth in complete and minimal media (Fig. 1.5). Prodiogiosin biosynthesis was found to be detectable after 8-12h of incubation at 27°C in liquid media, reaching a maximum after 24h. Although a positive regulator of prodigiosin biosynthesis (Rap) has been identified, this protein is also required for the expression of carbapenem in strain ATCC39006 (see section 1.3.4.2). It is therefore a reasonable assumption that Rap synthesis occurs from early in the growth phase onwards. So although Rap is directly or indirectly essential for pigment production, there must be additional physiological factor(s) which are also prerequisite(s) for prodigiosin biosynthetic gene expression.

As with many secondary metabolites, it has been found that extrinsic factors, such as culture conditions and varying nutrient supply have profound effects on prodigiosin production. A plethora of published information exists in this area. Some of these studies are described in the subsections below. Since growth rate is strongly influential in prodigiosin biosynthesis in *S. marcescens*, and factors affecting pigment production could not be measured quantitatively, a "non-proliferating-cells" (NPCs) system was developed to separate prodigiosin biosynthesis from cell proliferation. NPCs allow the effects of various factors on pigment biosynthesis to be determined without the complications of growth phase dependency. NPCs are made by growing *S. marcescens* for 48h at 38°C followed by washing and resuspending in saline at 27°C. These cells are capable of active metabolism but do not multiply (Williams *et al.*, 1971a).

Figure 1.4

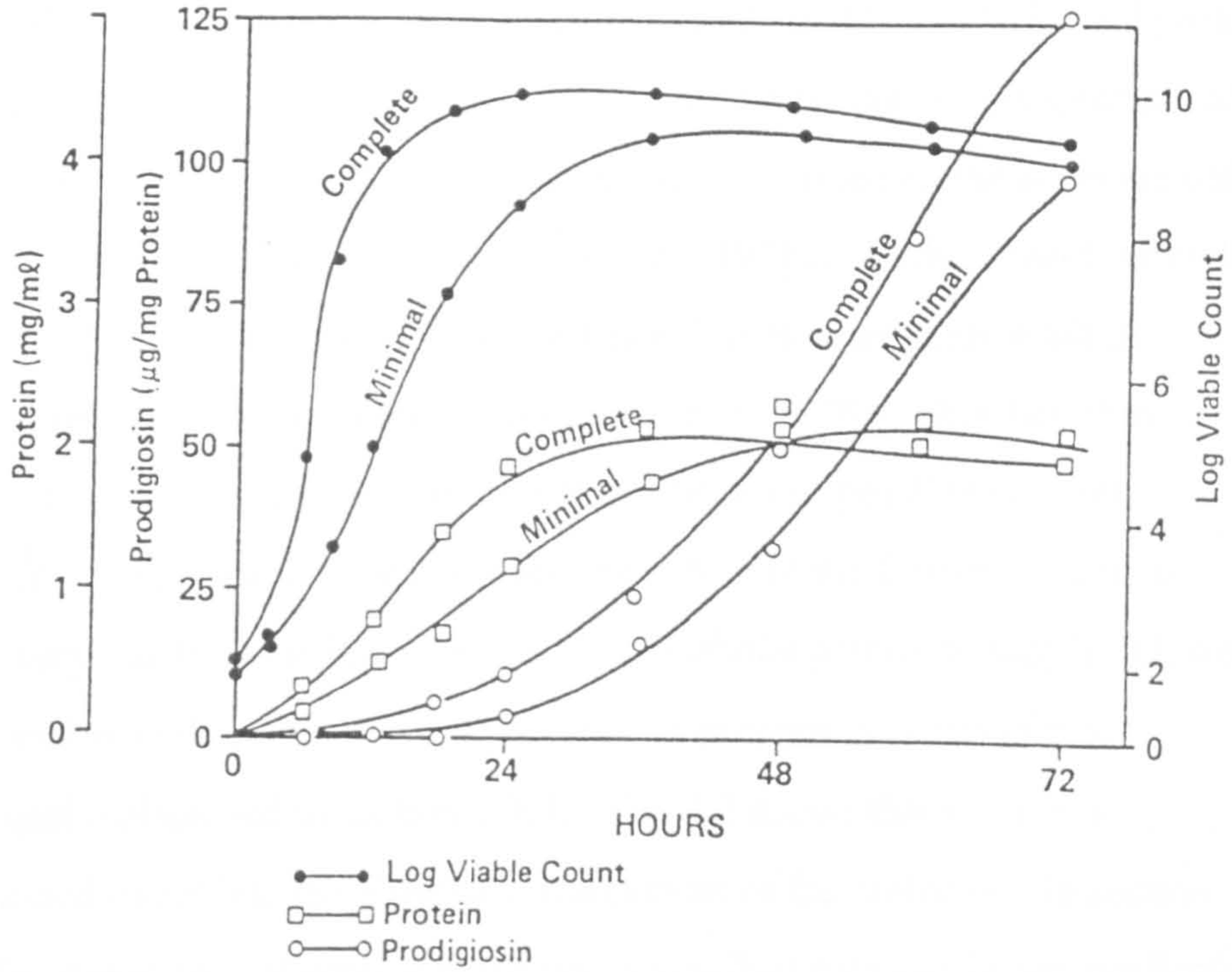


FIGURE 1.4 THE KINETICS OF PRODIGIOSIN BIOSYNTHESIS AS DETERMINED BY VIABLE COUNT AND PROTEIN, OF *Serratia marcescens*, INCUBATED AT 27°C

Bacteria were cultured in complete or minimal liquid media on a rotary shaker. Taken from Williams and Qadri (1980)

1.3.6.1 MEDIUM COMPOSITION, CARBON AND NITROGEN SOURCES

Pigment production in *S. marcescens* occurs readily in standard *E. coli* type complex media (Fig. 1.4), but is enhanced in a simple medium containing peptone and glycerol as sole nutrient sources (Morrison, 1966). Pigment biosynthesis in *S. marcescens* NPCs can be induced by the addition of certain amino acids to the culture medium. Addition of alanine, proline or histidine causes induction, but it was found in the same study that DNA, RNA and protein synthesis is also increased in the cells, implying that prodigiosin is synthesised secondarily to primary metabolic events in which supplied amino acids are utilised as carbon and nitrogen sources (Williams *et al*., 1971a). In this regard, virtually any supplied substrate which can be used as a C or N source can result in enhanced prodigiosin biosynthesis (Williams *et al*., 1976). This therefore indicates that synthesis of secondary metabolites is not possible without antecedent primary metabolism because regulation of the former is likely to depend very much on the latter, because of metabolic precursor supply. However, certain amino acids are known requirements as precursors for prodigiosin biosynthesis (discussed in section 1.3.3). Fig. 1.2 shows that proline is incorporated intact into the bipyrrolic component of the molecule. In section 1.3.5. the importance of proline in the analogous Red pathway in *Str. coelicolor* A3(2) was discussed, so clearly, this amino acid is essential in synthesis of these pigments. Proline is believed to have a dual role in this respect: it is both a direct precursor of prodigiosin and it is also an inducer of pre-required metabolic events which allow pigment formation to occur. *S. marcescens* **p**roline **ut**ilisation (Put) mutants, which are defective in the catabolic enzyme proline oxidase, were unable to synthesise prodigiosin in the NPC state when proline was supplied as the inducer (Scott *et al*., 1976); addition of methionine, together with other amino acids which could act a carbon or nitrogen sources to these overcame this block (Williams and Qadri, 1980), so proline oxidase is not required directly for pigment biosynthesis. Alanine can also play a dual role in pigment biosynthesis. It is suggested that alanine is the source of acetate moieties in prodigiosin biosynthesis (Wasserman *et al*., 1973), but it can also be utilised in primary metabolism. Histidine is not required as a direct precursor in prodigiosin biosynthesis and

therefore induces prodigiosin biosynthesis in NPCs only because it utilised in primary metabolism as a C or N source (Williams and Qadri, 1980). Aut (**A**lanine **ut**ilisation) and Put mutants are unable to catabolise alanine and proline respectively, and radiolabel tracer experiments showed that that greater amounts of these amino acids are inserted directly into prodigiosin in these mutants (Lim *et al*., 1976). Together, these studies showed that alanine, histidine and proline *catabolism* is not essential to prodigiosin biosynthesis unless these amino acids are supplied as sole carbon or nitrogen sources. Methionine has a synergistic role in prodigiosin biosynthesis (Qadri and Williams, 1973). Addition of methionine alone to NPCs did not induce pigment formation, but its addition with other effective amino acids resulted in a four-fold increase in the amount of prodiogiosin produced in comparison to adding the other amino acids singularly. Methionine also reduced the lag period prior to pigment formation in NPCs. Methionine is required for methylation of prodigiosin (section 1.3.3. and Fig. 1.2) but the researchers were unable to offer an explanation for its synergistic action in that study.

C and N requirements for pigment formation *in S. marcescens* have been found to be minimal, and prodigiosin is synthesised readily in a minimal medium containing glucose and ammonium salts. However, glucose repression has been demonstrated in the NPC system, where prodigiosin biosynthesis is abolished if glucose is added to culture medium at least 6h prior to the predicted onset of pigment production. The inhibition is not overcome by the addition of proline to the medium, until glucose supply is exhausted (Williams and Qadri, 1980). In a study in which glucose repression was investigated, it was reported that glucose repression in proline-induced cells was reversed by addition of a phosphodiesterase inhibitor, implying that repression of pigment biosynthesis is mediated via cyclic-AMP (cAMP) (Clemenys-Jewry, 1976). Catabolite repression via cAMP is a well characterised regulatory mechanism in enteric bacteria. It was also found in another study that the repressive effects of glucose might be mediated through inhibition of proline oxidase (Williams and Qadri, 1980), so the repression may well be mediated indirectly.

1.3.6.2 ANIONS AND CATIONS

Inorganic phosphate (P_i) has been found to have a profound effect on prodigiosin biosynthesis in both growing cells and NPCs (Williams and Qadri, 1980). ATP, ribose and P_i , but not adenine, were found to inhibit prodigiosin biosynthesis in growing cultures (Lawanson and Sholeye, 1976), although ATP hydrolysis by the cells was not demonstrated. The researchers concluded that the "endogenous content" of ribose and P_i in ATP caused the inhibition. It has been found that P_i inhibits synthesis of MAP and MBC in the prodigiosin biosynthetic pathway (Williams and Qadri, 1980). The mechanism of this inhibition is not known. Other researchers have found that increasing amounts of P_i proportionally diminish the burst of alkaline phosphatase activity which apparently occurs prior to the onset of prodigiosin biosynthesis (Witney *et al.*, 1977), and it was observed that this inhibition is nullified if P_i is added to the medium after the onset of pigment formation (Williams and Qadri, 1980). It is suggested that P_i inhibition is therefore mediated through inhibition of alkaline phosphatase, but this enzyme has not, to date, been assigned a role in pigment formation in *S. marcescens*.

Various effects of cations on pigment formation have been reported, including a possible requirements for K^+ and Fe^{2+} (reviewed in Williams and Qadri, 1980), but to date, these studies are purely phenomenological, with speculative explanations being offered for the effects observed. For example, Williams and Qadri (1980), cite one study in which sea salts were found to inhibit prodigiosin biosynthesis.

1.3.6.3 THE EFFECT OF TEMPERATURE

In *S. marcescens*, prodigiosin is produced in a relatively narrow temperature range, between 12 and 36°C (Williams and Qadri, 1980). Failure to form pigment at 37°C or higher is believed to result from inhibition of multiple (and unidentified) steps of the pathway, and it has been shown that temperatures of 40°C or higher inhibit condensation of MAP and MBC (Williams *et al.*, 1965), indicating that the condensing enzyme is temperature sensitive.

1.3.6.4 THE EFFECT OF pH

As mentioned previously, prodigiosin is a pH indicator, being red in acid and yellow in alkali. However, when pigmented strains of *S. marcescens* are grown on nutrient agar buffered at various pH values, pigmentation was found not to vary (Williams and Qadri, 1980). In a more recent study using NPCs, it was found that the optimum pH for pigment biosynthesis in buffered medium in these cells is 8.0-8.5 (Solé *et al.*, 1994). It has also been found that prodigiosin production in unbuffered medium was found to be the same regardless of initial pH, and pigmented cells of *S. marcescens* have a powerful buffering capacity and higher membrane proton conductance than non-pigmented mutants (Rius *et al.*, 1994).

1.3.6.5 THE EFFECT OF AERATION

It has been found that prodigiosin biosynthesis is optimal under conditions of high oxygenation, such as those achieved by culturing cells in baffled flasks with vigorous shaking (Heinemann *et al.*, 1970). Although *S. marcescens* is facultatively anaerobic, pigment production is inhibited under anaerobic conditions. Under such conditions, synthesis of MAP, MBC and their condensation is inhibited. It was also found in another study, that CO_2 is required by certain mutants, blocked in unidentified step(s) of the prodigiosin biosynthetic pathway (Williams and Qadri, 1980)

1.3.6.6 ANTIMICROBIAL AGENTS

The effects of agents which inhibit protein synthesis in *S. marcescens* have been studied in relation to pigment formation. It has been found that chloramphenicol, streptomycin and tetracycline can inhibit prodigiosin biosynthesis at concentrations which have little effect on cell division (Williams and Qadri, 1980). Synthesis of MBC was found to be more sensitive than MAP to streptomycin (Williams and Gott, 1964), and to be effective, inhibitors need to be added during lag period of bacterial growth (Siddiqui and Peterson, 1963). Also, agents which inhibit DNA synthesis, such as acraflavine and mitomycin C, disrupt

prodigiosin biosynthesis at concentrations which do not affect cell division (Williams and Qadri, 1980).

1.4 SECONDARY METABOLISM IN MICROORGANISMS

As stated previously, prodigiosin is a classical secondary metabolite of unknown function in *S. marcescens*. Although widely exploited by humans, secondary metabolism in biological systems is still the subject of much debate. In order to place prodigiosin biosynthesis in a wider biological context, aspects of secondary metabolism shall be discussed in this section.

1.4.1 DEFINITIONS OF SECONDARY METABOLISM

No single, unifying definition of secondary metabolism has been found, which perhaps reflects the biological diversity of the secondary metabolites themselves. Bu'lock (1961) considered secondary metabolism to be distinct from "general" (primary) metabolism by having no obvious function in producing organisms, and having restricted distribution. Williams *et al*., (1989) propose that secondary metabolites can be defined as "natural substances which do not play an explicit role in the internal economy of organisms that produce them." The latter definition encompasses the notion that secondary metabolites confer selective advantages on producing organisms. Vining (1990), defines secondary metabolites as "...substances [that] are not essential for growth and reproductive metabolism, [which are] usually distinctive products of particular groups of organisms, sometimes even of a single strain".

1.4.2 PROKAYOTIC SECONDARY METABOLITES

Plants, fungi and bacteria are capable of secondary metabolism. For relevance, only secondary metabolism in bacteria shall be considered in this and subsequent sections. Bacteria synthesise a wide variety of these compounds, ranging from complex antimicrobial agents to simple aliphatic acids (Vining, 1990). Examples of secondary metabolites are given in this section, but this list is by no means exhaustive.

Many of the known prokaryotic secondary metabolites are antibiotics. Actinomycetes, in particular the streptomycetes, are the most prolific group of antibiotic producers among bacteria, making three quarters of all known antibiotics (Demain, 1990). Actinomycete-derived antibiotics often have complex chemical structures, whereas those produced by unicellular bacteria are usually peptides or modified peptides. Exceptions to the latter include the prodigiosins, β-lactams and aminoglycosides. Taxonomic groups capable of synthesising antibiotic peptides includes enterics, pseudomonads and lactic acid bacteria (Vining, 1990). It has been proposed that antibiotics were originally formed via chemical (i.e., non-enzymatic) reactions and played an important evolutionary role in effecting and modulating primitive transcription and translation events by binding to receptor sites in primitive macromolecular templates made without enzymes; later on, these small molecules were replaced by polypeptides, but retained their abilities to bind to receptor sites in nucleic acids and proteins, and thus evolved to become antagonists in biological processes (Davies, 1990). Many of the commercially exploited antibiotics are antimicrobial agents. Antimicrobial agents produced by bacteria vary in specificity and are produced to confer competitive advantage in the environment. Some strains are able to produce agents active against other strains of the species, for example phenazine, produced by strains of *Pseudomonas phenazinium* (Messenger and Turner, 1981). Antimicrobial agents effective against other species of bacteria are numerous. Examples of these include carbapenem, produced by *S. marcescens* and *Erwinia carotovora* subsp. *carotovora* (Parker *et al*., 1982) and antibiotics produced by *Erwinia carotovora* subsp. *betavasculorum* and *Myxococcus xanthus*. A close correlation was found between the production of antibiotic by *Erw. carotovora* subsp. *betavasculorum in vitro*, and inhibition of the competing species *Erw. carotovora* subsp. *carotovora* in potato (Axelrod *et al*., 1988). Both species are phytopathogens. Myxobacteria, e.g. *Mxy. xanthus* feed on other bacteria at high cell densities, and fail to grow on *E. coli*, for example, unless $>10^7 ml^{-1}$ *Myx. xanthus* cells are present (Rosenberg and Varon, 1984). At high cell density, an antibiotic is produced, which is thought to be essential in killing and nutritional use of other bacteria (Demain, 1995). Prodigiosin also possess antibacterial

activity, and it has been found to be more effective against Gram-positive species that Gram-negative species (Grimont and Grimont, 1978). Bacteria also produce antibiotics which are effective against protozoa. Protozoa such as amoebae are able to predate bacteria, so certain bacterial species produce agents which are effective against them. *Pse. aeruginosa* produces pyocyanine, penicillic acid, phenazines and citrinin, which are able to kill protozoa (Demain, 1995). As mentioned previously (section 1.3.3), prodigiosin also has anti-protozoal activity, and it has been found that engulfment of pigmented strains of *S. marcescens* by protozoa results in encystment or death (Grimont and Grimont, 1978). Violacein, a pigmented secondary metabolite produced by *Chromobacteruim violaceum*, can also kill protozoa (Sivendra and Lo, 1975). Bacterial phytopathogens are able to produce secondary metabolites which are active against host tissues. Examples include wildfire toxin of *Pse. syringae* pv. *tabacum* and phaseolotoxin of *Pse. syringae* pv. *phaseolica* (Taylor *et al.*, 1972; Zhang *et al.*, 1993). Modes of action of these toxins can differ, but all benefit the producing organism. Production of tabtoxinine-β-lactam and coronatine by strains of *Pse. syringae* results in enhanced virulence, larger disease lesions, higher bacterial populations and increased survival of bacterial populations *in planta* (Mitchell, 1991). Herbicidins, produced by *Str. sagononensis*, are a group of nucleosides with selective phytotoxicity towards dicotyledenous plants (Arai *et al.*, 1976). Bacterial toxins produced by pathogens of higher animals are not generally regarded as secondary metabolites, although some bacteria are able to produce toxic antimetabolites (Vining, 1990).

Certain secondary metabolites function as mineral scavenging agents. One such group of substances is siderophores, which function in uptake, transport and solubilisation of ferric and ferrous ions. Many bacterial species have been shown to produce siderophores, including *Streptomyces* spp and *Pseudomonas* spp. These compounds can act as antibiotics by chelating iron to starve competitor species of the metal if competitors lack the means to take up the Fe-siderophore complex (Demain, 1995). Excretion of pseudobactin by *Pse. flourescens* and *Pse putida* is advantageous in the plant rhizosphere, where these species compete with phytopathogenic fungi (Venturi *et al.*, 1995). The other main group of mineral

scavenging agents is the ionophoric antibiotics, which function in the transport of alkali metal ions, e.g. macrotetralide antibiotics which enhance cell membrane permeability to K^+ (Demain, 1995). Production of a macrotetralide ionophore has been demonstrated to have survival advantage in *Str. griseus* in high Na^+ / low K^+ environments (Kanne and Zahner, 1976, cited in Demain, 1995). Other secondary metabolites are surface active lipopeptides, lipoproteins, glycolipids and lipopolysaccharides which may assist in colony growth or substrate assimilation. Examples include surfactin, produced by *Bacillus subtilis* (Nakano *et al*., 1992) and a surface-active agent produced by *Pse. aeruginosa* which assists in substrate breakdown when cells are cultured on hydrocarbons (Demain, 1995).

Some secondary metabolites act as signalling molecules in producing species. They include differentiation factors and autoregulator molecules. Antibiotics and other secondary metabolites are important in sporulation. In sporulating bacteria, antibiotic biosynthesis is often coincident with differentiation into spores, with the antibiotic being inhibitory to proliferation of vegetative cells of the producing organism. In bacilli, for example, peptide antibiotic production occurs in late-log phase of growth and early stages of sporulation (Demain, 1995). Antibiotic formation in actinomycetes is also coincident with differentiation, although most actinomycete secondary metabolites have no role in differentiation and function only as antibiotics (Vining, 1990). Antibiotic production is not obligatory for spore formation, for example mutants which cannot produce bacitracin (*Bacillus licheniformis*), myobacillin (*B. subtilis*), linear gramicidin (*Brevibacillus brevis*), tyrocidine (*Bre. brevis*), gramicidin-S (*Bre. brevis*), oxytetracycline (*Str. rimosus*), streptomycin (*Str. griseus*), and methylenomycin (*Str. coelicolor* A3(2)) have been isolated which are still able to sporulate (Demain and Piret, 1979). Secondary metabolites play a role in spore protection and inhibition of germination, for example gramicidin-S delays outgrowth of spores in *Bre. brevis* and *Str. viridochromogenes* spores contain an antibiotic that is associated with inhibition of germination (Demain, 1995).

1.4.3 AUTOREGULATOR MOLECULES AND BACTERIAL PHEROMONES

Autoregulator molecules with global functions in producing organisms are also considered to be secondary metabolites, and they are often classified as bacterial pheromones. This group of substances, produced by numerous bacteria, deserves special consideration in any discussion of the benefits of secondary metabolism. Autoregulator molecules which function as pheromones are often produced in cell-density dependent fashion in late-log to stationary phase and usually activate the production of virulence or differentiation factors and secondary metabolites in producing organisms. The view that bacteria are simple, relatively primitive organisms is now outdated. Obvious exceptions to view that prokaryotes are autonomous free-living unicellular organisms have been known for several decades, for example, the complex life cycles of streptomycetes and myxobacteria (Shapiro, 1988), and swarming behaviour in Gram-negative bacteria (Harshey, 1994). Increasingly though, it is being found that the individual bacterium is more analogous to a component cell of a multicellular organism; and bacterial cells exhibit features which show that temporal control of growth occurs during colony growth on agar, for example (Shapiro, 1988), which implies evolutionary sophistication that extends beyond the boundary of the individual cell.

A-factor (2-isocapryloyl-3-R-hydroxymethyl γ butyrolactone) is produced *by Str. griseus* and other streptomycetes (Hara *et al.*, 1983). A-factor activates streptomycin production, sporulation and yellow pigment production in *Str. griseus*, where it is required for transcription of the streptomycin activator *strR* (Vujaklija *et al.*, 1993). Other butyrolactones which are similar to A-factor have been reported, for example VB compound of *Str. virginae* (reviewed in Horinuchi and Beppu, 1994), and A-factor in *Str. cattleya*, which is required for production of thienamycin (Buchan *et al.*, 1994). In streptomycetes, it appears that A-factor type signals are species-specific and permit signalling to other cells within a colony by chemical diffusion. A-factor exerts its regulatory role by binding to and neutralising a receptor protein which, in the absence of A-factor, acts as a repressor-type regulator.

A-factor is structurally similar to the small diffusible acylated homoserine lactone (HSL) molecules produced by many other bacteria, most notably Gram-negative species. Quorum sensing is a term used to describe a system whereby bacteria are able to monitor their own population density (Fuqua *et al*., 1994). Bacteria in a colony produce a diffusible molecule (the pheromone) which accumulates in the environment during cell growth. The level of the pheromone increases concomitantly with cell density, until a critical threshold concentration is attained and the genes which require it for their activation are then induced. The number of bacterial species in which HSL-dependent quorum sensing occurs is constantly increasing and is therefore too numerous to discuss here. They are extensively reviewed in Hussain *et al*., (1998), Gray (1997), Salmond *et al*. (1995), and Fuqua (1994). However, as mentioned previously (section 1.3.4.2), OHHL controls carbapenem and virulence factor production in *Erw. carotovora* subsp. *carotovora*, but an OHHL-like molecule has not been identified in *S. marcescens* strain ATCC39006, although other strains of *S. marcescens* do produce HSLs. A variety of processes are known to be regulated in a cell-density- or growth-phase-dependent manner in Gram-positive bacteria. Examples of such quorum-sensing systems are the development of genetic competence in *B. subtilis* and *Streptococcus pneumoniae*, the virulence response *in Staphylococcus aureus*, and the production of antimicrobial peptides by several species of Gram-positive bacteria, including lactic acid bacteria. Cell-density-dependent regulatory modes in these systems use signalling molecules which are post-translationally processed peptides, secreted by a dedicated ATP-binding-cassette exporter. This secreted peptide pheromone functions as the input signal for a specific sensor component of a two-component signal-transduction system (Kleerebezum *et al*., 1997).

1.4.4 WHY DOES SECONDARY METABOLISM EXIST?

As alluded to in section 1.4.1, secondary metabolism is not easily defined, nor categorised, in a simple and unified way. There has been much debate about why living organisms have the ability to perform non-essential complex syntheses when evolution tends to favour energetic economy. Organisms

that are capable of secondary metabolism have been said to possess "*...a playground for games in which a reasonable level of low-cost inventive evolution is tolerated*", (Zahner, 1979). Although examples of secondary metabolites with biological functions were given in sections 1.4.2 and 1.4.3, many secondary metabolites have not been assigned functions in producing organisms. It is suggested that some secondary metabolites only appear to have no benefit to producing organisms because of lack of suitable assay procedures to detect their activity (Kell *et al.*, 1995).

Secondary metabolic pathways were once regarded as spin-offs from primary metabolism, or even as the aberrant results of systems that have suffered a breakdown of metabolic control, or simply the results of some sort of genetic and metabolic idiosyncrasies. Vining (1990 and 1992), argues that the sophistication of secondary metabolite biosynthesis, its regulation and genetic organisation argue for it being an activity of greater benefit to the producing organism. Secondary metabolism often involves novel pathways which are distinct from those of primary metabolism. A good example here is prodigiosin itself: the pyrrole rings of prodigiosin are synthesised from amino acids and acetate (section 1.3.3), unlike, for example, the pyrrole rings of porphyrins involved in primary metabolic pathways, which are synthesised through the succinate-glycine cycle. Another example is the biosynthesis of polyketides. Polyketides are a widely distributed and highly diverse group of secondary metabolites which bear obvious similarities to the general pathway for fatty acid biosynthesis in biological systems. The principal differences lie in the larger variety of acyl precursors incorporated, and the partial or complete absence of reactions that reduce the β-keto group formed after each chain extension in polyketide pathways. As a result, polyketides are structurally more diverse than fatty acids because their intermediates undergo a variety of condensations and cyclisations to yield complex aromatic or macrocyclic products. The acyl-carrier proteins (ACPs) in polyketide pathways are distinct from those of fatty acid metabolism and their genes are usually located in polyketide pathway loci (Vining, 1992). Polyketides are synthesised by large multimeric and multifunctional synthase complexes, and their encoding genes possess similarities to those of fatty acid synthase complexes (Hopwood and

Sherman, 1990). DNA sequence similarities indicate that polyketide synthases share a common evolutionary origin with fatty acid synthases (Bibb *et al.*, 1989; Sherman *et al.*, 1989), but further comparison in certain systems does not support the simple hypothesis that polyketide synthases evolved from fatty acid synthases in polyketide producing organisms (Revill and Leadlay, 1991; Donadio *et al.*, 1991).

The concept that microorganisms have evolved the ability to produce secondary metabolites because they create a competitive advantage is supported by several arguments (Stone and Williams, 1992): (i) only organisms lacking an immune system are prolific producers of these compounds; (ii) these metabolites have sophisticated chemistry and modes of action, and are produced by energetically expensive pathways; (iii) environmental isolates synthesise natural products, most of which have physiological functions; (iv) clustering of secondary metabolite biosynthetic genes is commonplace, and devoid of non-functional genes; (v) secondary metabolite gene clusters often contain resistance and regulatory genes.

Various hypotheses have been put forward to explain why secondary metabolism should have evolved, and the arguments for and against each of these hypotheses can be found in Vining (1990). Many of these hypotheses do not place a great deal of importance on the secondary metabolite *per se*, but rather on the possible reasons why the pathways of secondary metabolism should exist in living organisms, which is perhaps a reflection of the difficulty biologists have had in perceiving a single theme common to every example of secondary metabolism. For example, the shunt hypothesis proposes that secondary metabolic pathways are routes for dissipating excess carbon and energy under physiologically stressful conditions. Through these pathways, an accumulated intermediate metabolite is modified by a succession of enzymatic reactions which result in formation of a complex product whose biological activities are "accidental" (Vining, 1990). As mentioned previously, a variation on this theme is proposed to explain the production of Red in *Str. coelicolor* A3(2) (see section 1.3.5.2).

1.4.5.1 GENES AND GENETIC ORGANISATION OF PROKARYOTIC SECONDARY METABOLITE BIOSYNTHETIC PATHWAYS

Secondary metabolites are derived from pathways which are unique and distinct from those of primary metabolism. Pathways are often long and complex and require catalysis by dedicated enzymes. Regulation of genes of secondary metabolism involves control of timing and level of biosynthetic gene expression. These control mechanisms have been found to be well integrated in the physiology of the producing organism. Vining, (1992) argues, if genetic organisation of DNA encoding secondary metabolites were not maintained by selection, chance mutations that occasionally arise, even those that pose a low metabolic load, should have eventually led to elimination of these genes. The presence of gene clusters encoding secondary metabolism therefore strongly implies positive selection.

Clustering of genes for secondary metabolite formation is a common phenomenon, and is particularly well studied in *Streptomyces* spp. among the prokaryotes. Since transcription and translation are tightly coupled in prokaryotes, ordering of structural genes allows each enzyme formed by coupled transcription-translation to act on the product of the previous enzyme of the pathway, so the enzyme need not find its substrate by mere random collision. Examples of gene clusters encoding secondary metabolites are numerous and include the following (Martín, 1992): tyrocidines, gramicidin-S and bacitracins, which are peptide antibiotics produced by *Bacillus* spp; bialaphos, a herbicide and antibiotic produced by *Str. hygroscpoicus* and *Str. viridochromogenes*; erythromycin produced by *Saccharopolyspora erythrea*; tetracycline, produced by *Str. rimosus* and tetracenomycin C, produced by *Str. glaucescens* and β-lactam antibiotics produced by *S. marcescens* and *Erw. carotovora* subsp. *carotovora*. The situation would be predicted to be complex in producers of secondary metabolites which utilise branched pathways, e.g. prodigiosin and undecylprodigiosin biosynthesis. A more complex genetic organisation is apparent in *Str. fradiae*, which produces tylosin, a macrolide antibiotic. Nine tylosin biosynthetic genes and one resistance gene map to a 42 Kb stretch of chromosomal DNA. Genes were found to be ordered, and are divided into so-called early and late genes, with the resistance

gene being located at one end. Organisation into subclusters appears to exist within DNA of tylosin late genes. Other early genes are not present in this cluster, including the gene encoding tylactone, a tylosin intermediate (Fishman *et al.*, 1987). Efficiency of secondary metabolite production is also increased by the production of multi-enzyme complexes to catalyse sequential steps in the biosynthetic pathway, for example the aforementioned polyketides synthases (section 1.4.3). This arrangement is also used in biosynthesis of gramicidin S, tyrocidines and bacitracins in *Bacillus* spp.(Martín and Liras, 1989).

In common with most evolutionary events, it is likely that the events through which secondary metabolism arose, have only infrequently resulted in a positive outcome which conferred a selective advantage on the host organism. Evidence suggests that genes encoding secondary metabolites show much weaker homology to genes of related primary pathways in the same organism, than to those of related primary and secondary pathway genes in other organisms, which implies that gene transfer has been an important factor in the spread of secondary metabolism (Vining, 1992). Also, secondary metabolism may have been the route by which biochemistry advanced after the "housekeeping" primary metabolic pathways had evolved. Useful functions which occasionally arose through this might have been acquired horizontally by other species and then adapted for individual advantage. This might explain, therefore, why *S. marcescens* and *Str. coelicolor* A3(2), which are not closely related, both have the ability to synthesise prodigiosins which differ from each other only in the length of a hydrocarbon side chain.

1.4.5.2 EXPRESSION OF GENES ENCODING SECONDARY METABOLITES

Studies have shown that certain mechanisms which regulate gene expression in prokaryotes may be involved in expression of secondary metabolite genes, which are generally induced in late-log to stationary phases of growth. The production of alternative sigma factors, or other ancillary proteins that alter the specificity of RNA polymerase is a mechanism by which gene expression can be controlled. In bacteria, control of gene expression at the transcriptional level is

afforded by "reprogramming" of core RNA polymerase with so-called alternative (non σ^{70}) sigma factors, which allows recognition of non-standard promoters. This allows certain regulons to be induced in certain specific conditions. Several classes of alternative sigma factors have been identified, and various studies have enabled promoter consensus sequences to be deduced for many of them (elaborated upon in section 3.4.3.2). The main alternative sigma factors are as follows: Sigma-54 (RpoN) is involved in the nitrogen starvation response in *E. coli* (Helmann and Chamberlin, 1988); it is also involved in virulence factor expression in *Pse. aeruginosa* (Ishimoto and Lory, 1989; Totten *et al.*, 1990). Sigma-32 (RpoH) is a global regulator of the heat shock response in bacteria. Some of the genes induced in the presence of RpoH are also induced under other stress conditions, for example, exposure to hydrogen peroxide, ethanol, bacteriophage infection, or mis-folded proteins (Dorman, 1993). Sigma-24 (RpoF) regulates genes induced in the "extreme" heat shock response (Ishihama, 1993). However, the second most important sigma factor in bacteria is Sigma-38 (RpoS). This has been characterised in many studies of diverse cellular processes as the principal stationary phase sigma factor. RpoS-induced genes include those which allow cell survival without cell division in stationary phase, for example, in *E. coli*, the *bolA*, morphogene (Aldea *et al.*, 1989; Lange and Hengge-Aronis, 1991), *katE*, encoding H_2O_2-resistance (Mulvey *et al.*, 1990), *dps* encoding DNA protection (Almirón *et al.*, 1992), *osmB* encoding lipoprotein (Jung *et al.*, 1990) and *osmY* encoding periplasmic protein (Lange *et al.*, 1993). Not surprisingly, genes involved in secondary metabolism have also been found to require RpoS. Examples include *mcc*, required for microcin C7 expression in *E. coli* (Diáz-Guerra *et al.*, 1989; Moreno *et al.*, 1995), and very recently, it has been found that expression of *rsmA*, encoding a global regulator of secondary metabolism and virulence in *Erw. carotovora* subsp. *carotovora*, is controlled by RpoS (Mukherjee *et al.*, 1998). An RpoS homologue has been recently found to regulate HSL-dependent autoinduction in the phytopathogen *Ralstonia* (*Pseudomonas*) *solanacearum* (Flavier *et al.*, 1998). Several promoters of antibiotic biosynthesis and resistance genes which have been sequenced, including those of streptomycetes. Their genetic organisation can be complex and they do not

resemble either *E. coli* or *B.subtilis* consensus promoter sequences (reviewed in Matín and Liras, 1989). Other alternative sigma factors have been well studied in *B. subtilis*, and they are required for differentiation during spore formation (reviewed in Helmann and Chamberlin, 1988).

A further example of promoter complexity in secondary metabolite operons is found in DNA encoding microcin B17. This is a low molecular weight, bacteriocin-like compound which is a classical secondary metabolite in *E. coli*. Growth phase induction of microcin B17 genes has been studied, and it was found that the major promoter of the microcin B17 biosynthetic operon is characteristic of a class of promoters called "gearbox promoters" (Aldea *et al.*, 1990). These promoters have -35 region homologous to that recognised by *E. coli* Sigma-70, but possess differing -10 regions (Vicente *et al.*, 1991). Deletion of the -10 site disrupts growth phase control (Aldea *et al.*, 1990). Expression from gearbox promoters has been shown to inversely proportional to growth rate by promoter fusion studies. Transcription regulation from these promoters may involve alternative sigma factors, for example, expression of the *ftsQ* gene involved in cell division in *E. coli* possesses a gearbox promoter which may be RpoS-dependent (Sitnikov *et al.*, 1996). The different factors discussed in this section serve to highlight that control of secondary metabolite gene expression is very complex, involving many different input signals, and is regulated in a multi-layered manner.

1.5 WHY STUDY PIGMENT PRODUCTION IN *Serratia marcescens*?

Although secondary metabolites, most notably antibiotics, are heavily exploited by humans, there is clearly much that remains to be learnt about the processes leading to their production. Understanding the expression and regulation of genes of secondary metabolism is important in establishing how they evolved in comparison to genes of primary metabolism. Also, any research which increases in our knowledge and understanding of secondary metabolism has potential relevance to industry and biotechnology. Due to taxonomic relatedness, *S. marcescens* strain ATCC39006 is sensitive to a repertoire of molecular genetic techniques which is reasonably comparable to *Erw. carotovora* subsp. *carotovora*. In this sense, the fact that it synthesises prodigiosin is both interesting in itself (as

the number of publications concerning this phenomenon attests to), but it is also an invaluable resource for investigating secondary metabolism as a model system. Given the technologies which are currently available, prodigiosin biosynthesis can be investigated at the most fundamental, i.e. genetic, level. As shall be explained in subsequent chapters as appropriate, strain ATCC39006 can be made sensitive to infection by coliphage λ, which allows easy strain construction; a *Serratia*-specific generalised transducing phage, ΦOT8 is available in this laboratory, which also facilitates strain construction. Since *S. marcescens* is a Gram-negative enteric species, it is relatively fast-growing and non-fastidious, so culturing the strain is not complicated by a complex life cycle. It is therefore more genetically amenable than *Str. coelicolor* A3(2), in which undecylprodigiosin has been the subject of nearly two decades of genetic analysis.

1.6 AIMS OF THIS STUDY

In 1980, Williams and Qadri stated, "Only speculations and anecdotal reports exist about the genetics of prodigiosin biosynthesis in *S. marcescens*...this information is paramount to a complete understanding of pigmentation and the function of prodigiosin in *S. marcescens*." Little progress has been made elucidating the genetics of prodigiosin biosynthesis since this statement was made. A major breakthrough in genetic characterisation of prodigiosin biosynthesis was made in this laboratory when prodigiosin biosynthetic genes were fully expressed in a heterologous host (section 1.3.4.2). For the reasons given in section 1.5 above, the overall aim of this study was to analyse prodigiosin biosynthesis at the genetic level in order to further our understanding of secondary metabolism in this species. A more detailed breakdown of this aim is given below:

(i) Sequencing of the prodigiosin biosynthetic gene cluster.

As prodigiosin biosynthesis could be reconstituted in *Erw. carotovora* subsp. *carotovora* by transformation with one complementing cosmid (pNRT104, see section 1.3.4.2), it was assumed that all the necessary biosynthetic genes are present in a cluster (named the *pig* cluster) within the 30-35 Kb of *S. marcescens* DNA carried on this cosmid. This assumption was supported by

the fact that the *red* cluster in *Str. coelicolor* A3(2) is at least 30 Kb in size, and encodes at least 18 Red proteins. *S. marcescens* DNA was subcloned from pNRT104 by A. Cox and smaller subclones were generated for sequencing. Sequencing of *pig* cluster DNA was to be divided between myself and A. Cox. Putative protein encoded in the *pig* cluster would then be expressed *in vivo*.

(ii) Generation of mutants that are defective in prodigiosin biosynthesis and carry instead reporter gene fusions.

While sequencing was in progress, random transposon mutagenesis strategies would be used to generate prodigiosin biosynthetic mutants. The aim was to produce a bank of mutants which could be assigned to corresponding open reading frames when sequence data became available. Use of transposons carrying reporter genes would allow construction of gene fusion strains. Different reporter genes strategies were to be attempted.

(iii) Use of reporter gene fusion strains in physiological experiments.

In the final part of the investigation, and time permitting, gene fusion strains constructed as described in (ii) above would be used as sensitive assays, (c.f. production of prodigiosin itself), to investigate the effects of some of the extrinsic factors reported to influence prodigiosin biosynthesis in this species, on individual *pig* gene induction. Expression of *pig* genes would be assayed indirectly, in relation to growth phase.

CHAPTER 2

MATERIALS AND METHODS

2.1 BACTERIAL STRAINS, PLASMIDS, COSMIDS AND BACTERIOPHAGES

STRAIN	CHARACTERISTICS	MARKERS	SOURCE and REFERENCE
Escherichia coli **strains:**			
DH5α	*supE44, ΔlacU169* (Φ80*dlacΔM15*) *hsdR17 recA1 endA1 gryA69 thi-1 relA1*		Lab stock Sambrook *et al.*, 1989
TG1	F' (*lac-pro*), *supE, thi*, F'*traD36, proAB*, *lacI*^{q}, ZDM15, *ecoK*		Lab stock Carter *et al.*, 1985
ESS	β-lactam super sensitive		L. Porter (provided by Smith-Kline Beecham) Bainton *et al.*, 1992
LE392	*F', hsdR514*, ($r_k^- m_k^+$), *supE44, supF58, lacY1, galK2, galT22, metB1, trp55*		Lab stock, Sambrook *et al.*, 1989
K38	K-12 (wild-type), HfrC, λ^+		G. Bosgelmez, Tabor and Richardson, 1985
Serratia marcescens **strains:**			
ATCC39006	Wild-type, Pig^+, Car^+		Lab stock, Bycroft *et al.*, 1987
SCAR952	ATCC39006::Tn*5*, $Restriction^-$, Pig^+, Car^+, Lac^+	Kn^R	A. Cox, unpublished
SCAR1310124	$Restriction^-$, Pig^+, Car^+, Lac^+		A. Cox, unpublished
SLAC1	$Restriction^-$, Pig^+, Car^+, Lac^-		This study
NTM2	*rap*::*kan*^{R} marker exchange mutant Pig^-, Car^-	Kn^R	N. Thomson, Thomson, 1996
SYA100	SCAR1310124::Tn*phoA*'-2, Pig^-, Car^+	Tc^R	This study
SYA101	SCAR1310124::Tn*phoA*'-2, Pig^-, Car^+	Tc^R	This study
SYA102	SCAR1310124::Tn*phoA*'-2, Pig^-, Car^+	Tc^R	This study
SYA105	SCAR1310124::Tn*phoA*'-2, Pig^-, Car^+	Tc^R	This study
SYA111	SLAC1::Tn*phoA*'-2, Pig^-, Car^+, Lac^+	Tc^R	This study

(Table 2.1 continued...)

SYA113	SLAC1::Tn*phoA*'-2, Pig^-, Car^+ Lac^+	Tc^R	This study
SYA300	SCAR1310124::Tn*phoA*'-2, Pig^+ (orange), Car^+	Tc^R	This study
SYA500	SCAR1310124::Tn*phoA*'-2, Pig^{++}(hyperpigmented), Car^+	Tc^R	This study
SYA12	SCAR952::Tn*blaM*, Pig^-, Car^+	Kn^R, Sp^R	This study
SYA14	SCAR952::Tn*blaM*, Pig^-, Car^+	Kn^R, Sp^R	This study
SYA20	SCAR952::Tn*blaM*, Pig^-, Car^+	Kn^R, Sp^R	This study
SYA22	SCAR952::Tn*blaM*, Pig^-, Car^+	Kn^R, Sp^R	This study
SYA24	SCAR952::Tn*blaM*, Pig^-, Car^+	Kn^R, Sp^R	This study
SYA25	SCAR952::Tn*blaM*, Pig^-, Car^+	Kn^R, Sp^R	This study
SYA30	SCAR952::Tn*blaM*, Pig^+ (orange), Car^+	Kn^R, Sp^R	This study
SYA50	SCAR952::Tn*blaM*, Pig^{++} (hyperpigmented), Car^+	Kn^R, Sp^R	This study
SYA1001	SYA100::SLAC1, Pig^-, Car^+, Lac^-	Tc^R	This study
SYA1011	SYA101::SLAC1, Pig^-, Car^+, Lac^+	Tc^R	This study
SYA1021	SYA102::SLAC1, Pig^-, Car^+, Lac^+	Tc^R	This study
SYA1051	SYA105::SLAC1, Pig^-, Car^+, Lac^-	Tc^R	This study
NTSA1	SLAC1::SYA1011::*rap*::Kn Pig^-, Car^-,Lac^-	Tc^R, Kn^R	This study

TABLE 2.2 PLASMIDS, COSMIDS AND BACTERIOPHAGES

PLASMID / COSMID	CHARACTERISTICS	PHENOTYPE	SOURCE and REFERENCE
pSF6	low-copy cosmid cloning vector, *cos*, *mob*	Sp^R, SmR	N. Thomson Selveraij *et al.*, 1984
pDAH330	High-copy cloning vector (pIC19R derivative)	Cm^R	A. Cox. unpublished
pBluescript II KS	multi-copy cloning vector, *lacZ*	Ap^R	Stratagene Ltd Yanisch-Perron *et al.*, 1985
pACYC177	low-copy cloning vector	Kn^R, Ap^R	NEB (UK) Ltd Chang and Cohen, 1978
pNRT104	pSF6 carrying *S. marcescens* prodigiosin biosynthetic cluster	Sp^R, Sm^R	N. Thomson Thomson, 1996
pM245	pDAH330 carrying *S. marcescens* prodigiosin biosynthetic cluster	Cm^R	A. Cox unpublished
pTROY9	pLAFRB *malK*::IS-3, $lamB^+$	Tc^R	de Vries *et al.*, 1984
pMUT13	pBR322::$lamB^+$	Kn^R	J. Thomas Clement *et al.*, 1982
pGP1-2	T7 polymerase	Kn^R	Tabor and Richardson, 1985

(Table 2.2 continued...)

pT7-6	T7Φ10 promoter	Ap^R	Tabor and Richardson, 1985
pTONB08M	pDAH330 carrying 0.8 Kb prodigiosin biosynthetic cluster *Bam*HI subclone	Cm^R	A. Cox unpublished
pTONB10M	pDAH330 carrying 1.0 Kb prodigiosin biosynthetic cluster *Bam*HI subclone	Cm^R	A. Cox unpublished
pTONB20M	pDAH330 carrying 2.0 Kb prodigiosin biosynthetic cluster *Bam*HI subclone	Cm^R	A. Cox unpublished
pTONB32M	pDAH330 carrying 3.2 Kb prodigiosin biosynthetic cluster *Bam*HI subclone	Cm^R	A. Cox unpublished
pTONB54M	pDAH330 carrying 5.4 Kb prodigiosin biosynthetic cluster *Bam*HI subclone	Cm^R	A. Cox unpublished
pTONB64M	pDAH330 carrying 6.4 Kb prodigiosin biosynthetic cluster *Bam*HI subclone	Cm^R	A. Cox unpublished
pTONB65M	pDAH330 carrying 6.5 Kb prodigiosin biosynthetic cluster *Bam*HI subclone	Cm^R	A. Cox unpublished
pBSB27	pBluescript II KS carrying 2.7 Kb prodigiosin biosynthetic cluster *Pst*I-*Kpn*I subclone	Ap^R	This study
pBSB26	pBluescript II KS carrying 2.7 Kb prodigiosin biosynthetic cluster *Kpn*I-*Bam*HI subclone	Ap^R	This study
pBSB32	pBluescript II KS carrying 2.7 Kb prodigiosin biosynthetic cluster *Kpn*I-*Bam*HI subclone	Ap^R	This study
pT76-PB	pT7-6 carrying Kb prodigiosin biosynthetic *Pst*I-*Bam*HI subclone	Ap^R	This study
pT76-PE	pT7-6 carrying Kb prodigiosin biosynthetic *Bam*HI-*Eco*RI subclone	Ap^R	This study
pACYC177/ 100	pACYC177 carrying Tn*phoA*'-2 insertion in *S. marcescens* chromosomal DNA from strain SYA100 on a *Bam*HI fragment	Kn^R, Tc^R	This study
pACYC177/ 101	pACYC177 carrying Tn*phoA*'-2 insertion in *S. marcescens* chromosomal DNA from strain SYA101 on a *Bam*HI fragment	Kn^R, Tc^R	This study
pACYC177/ 102	pACYC177 carrying Tn*phoA*'-2 insertion in *S. marcescens* chromosomal DNA from strain SYA102 on a *Bam*HI fragment	Kn^R, Tc^R	This study

(Table 2.2 continued...)

pACYA177/ 111	pACYC177 carrying Tn*phoA*'-2 insertion in *S. marcescens* chromosomal DNA from strain SYA111 on a *Bam*HI fragment	Kn^R, Tc^R	This study
pACYC177/ 113	pACYC177 carrying Tn*phoA*'-2 insertion in *S. marcescens* chromosomal DNA from strain SYA113 on a *Bam*HI fragment	Kn^R, Tc^R	This study
PHAGE			
M13 mp18	Sequencing vector, *lacZ*$^+$		Amersham Int. Messing and Vieira, 1982
M13 mp19	Sequencing vector, *lacZ*$^+$		Amersham Int. Messing and Vieira, 1982
λ::Tn*blaM*	coliphage carrying transposon		M. Sebaihia, Taddyon and Broome-Smith, 1992
λ::Tn*phoA*'-2	coliphage carrying transposon		M. Sebaihia, Wilmes-Reisenberg and Wanner, 1992
ΦOT8	*S. marcescens* generalised transducing phage		S. Bentley Orme and Thomson, unpublished; Cox (1996)
ΦOT8-SYA100	ΦOT8 propagated on *S. marcescens* strain SYA100		This Study
ΦOT8-SYA101	ΦOT8 propagated on *S. marcescens* strain SYA101		This study
ΦOT8-SYA102	ΦOT8 propagated on *S. marcescens* strain SYA102		This study
ΦOT8-SYA105	ΦOT8 propagated on *S. marcescens* strain SYA105		This study
ΦOT8-SYA1011	ΦOT8 propagated on *S. marcescens* strain SYA1011		This study

2.2 GROWTH AND MAINTENANCE OF BACTERIAL STRAINS AND BACTERIOPHAGES

2.2.1 MEDIA AND SOLUTIONS

Media and solutions used for routine growth and maintenance of bacterial and phage stocks are listed in Table 2.3. Media and solutions were sterilised by autoclaving at 121°C for 15 min and 120 psi pressure. Items shown in brackets were added to the growth medium after autoclaving from sterile stocks. Media were solidified by the addition of 1.6% (w/v) Bacto agar. Antibiotics and media supplements listed in Table 2.3 were prepared in ELGA water, filter sterilised unless

otherwise stated, and added to the growth medium after autoclaving and cooling of the medium.

Serratia marcescens and *Escherichia coli* strains were maintained on NB Agar (NBA) plates, at 4°C, for up to 3 months. Stocks for long term storage were cultured in liquid culture, under the appropriate conditions, and frozen at -70°C in 2x Freezing Medium. All phage stocks were stored at 4°C over a few drops of chloroform.

TABLE 2.3 MEDIA AND REAGENT CONSTITUENTS

MEDIUM OR SOLUTION	CONSTITUENTS PER LITRE
Nutrient broth (NB)	13g Oxoid nutrient broth
NB Agar (NBA)	NB, 16g Bacto agar (Difco)
Luria Bertani broth (LB)	10g Bacto tryptone 5g Bacto yeast extract 5g NaCl
LB agar (LBA)	LB, 16g Bacto agar
DD Agar (DDA)	20g Bacto Tryptone 8g NaCl (10 ml 1M Mg SO_4)
Minimal Medium (MM)	(20 ml 50x $PO_{4)}$ (10 ml 10% w/v $(NH_4)_2SO_4$) (10 ml 20% w/v C-source) as stated
MM agar (MMA)	MM, 16g Bacto agar (Difco)
Top agar for phage titration and carbapenem assay lawn	NB or DD as appropriate, solidified with 0.7% Bacto agar
Top agar for phage lysates	NB or DD, as appropriate, solidified with 0.3% Bacto agar
E. coli TG1 maintenance medium	16g Bacto agar (5 ml 40% w/v glucose) (20 ml 50x PO_4) (10 ml 10% w/v $(NH_4)_2SO_4$) (10 ml 1% w/v $MgSO_4.7H_2O$) (8 ml 5mgml^{-1} thiamine)
1.7x M9 Medium	10.2g Na_2HPO_4 5.1g KH_2PO_4 0.85g NaCl 1.7g NH_4Cl [pH7.4] (3.4 ml $MgSO_4$) (17 ml 20% w/v glucose) (170µl IM $CaCl_2$)
2X Freezing Medium	126g K_2HPO_4 0.9g trisodium citrate 0.18g $MgSO_4.7H_2O$ 3.6gKH_2PO_4 1.8g $(NH_4)_2SO_4$ 88 ml glycerol
Phage (Φ) Buffer	10mM Tris base 10mM $MgSO_4$ 0.01% w/v gelatine (pH 7.4)

TABLE 2.3 Continued…

Glucose-TE for small-scale plasmid preparation	20 ml 1M glucose 25 ml 1M Tris.Cl pH8.0 200 ml 1M EDTA pH8.0 745 ml dH_2O
SDS/NaOH for small-scale plasmid preparation	20 ml 10M NaOH 100 ml 10% SDS 880 ml dH_2O
K-Acetate for small-scale plasmid preparation	600 ml potassium acetate 115 ml glacial acetic acid 285 ml dH_2O
Treated chloroform for phage work	4-6 spatulas of Na_2CO_3 (anhydrous), shaken, settled and decanted and stored in the dark
50x Phosphate Buffer	350g K_2HPO_4 100g KH_2PO_4 (pH 6.9-7.1)
Z Buffer for β-galactosidase assay	8.52g Na_2HPO_4 6.24g $NaH_2PO_4.2H_2O$ 0.75g KCl 0.25g $MgSO_4.7H_2O$ 2.7 ml β-mecarptoethanol (pH 7.0)
10x Tris-borate-EDTA (TBE) Buffer	108g Tris base 55g Boric acid 9.3g EDTA
Tris-EDTA (TE) Buffer	10mM Tris base 1mM EDTA (Na^+ salt) Adjusted to required pH with HCl
DNA Loading Buffer	0.25% w/v bromophenol blue 25% w/v Ficoll Make up volume with dH_2O
40% Acrylamide stock	38% w/v acrylamide 2% w/v *NN*-methylene-*bis*acrylamide filtered through 2% w/v Amberlite resin, stored at 4°C in the dark
Denaturation Solution for Southern blotting	87g NaCl 20g NaOH
20x SSC	175g NaCl 88g trisodium Citrate

(Table 2.3 continued...)

Neutralisation Solution for Southern blotting	87g NaCl 121g Tris-HCl (pH 8.0)
DIG Prehybridisation Solution (Digoxigenin labelling)	25 ml 20x SSC 1.5g Blocking Reagent (from DIG Labelling Kit, Boehringer Mannheim) 1 ml 10% w/v N-Lauroyl Sarcosine 74.8 ml sdH_2O Stored at -20°C
Wash I (2x SSC)	100 ml 20x SSC 10 ml 10% w/v Sodium Doecyl Sulphate (SDS) 890 ml dH_2O
Wash II (0.5x SSC)	10 ml 20x SSC 10 ml 10% w/v SDS 980 ml dH_2O
DIG Buffer I (Malate Buffer)	11.61g Malic Acid (0.1M) 6.43g NaCl (0.15M)
DIG Buffer II	10 ml (10% w/v Blocking Reagent) 90 ml DIG Buffer I
DIG Buffer III	10 ml (1M Tris base) (0.1M) 10 ml (1M NaCl) (0.1M) 5 ml (1M $MgCl_2$) (0.05)
EMS Buffer A	60mM K_2HPO_4 33mM KH_2PO_4 7.6mM $(NH_4)2SO_4$ 1.7mM trisodium citrate
PEG precipitation solution	20g polyethylene glycol mw 6000 14.6g NaCl Volume adjusted to 1L with dH_2O
SDS-PAGE Acrylamide stock	30% w/v acrylamide 0.8% w/v bisacrylamide Filter, store at 4°C
SDS-PAGE 4x Lower buffer	1.5M Tris.Cl (pH 8.8) 0.4% w/v SDS 0.1% TEMED (pH 9.0)
SDS-PAGE 4x Upper Buffer	0.5M Tris.Cl (pH 6.8) 0.4% w/v SDS 0.2% TEMED (pH 6.8)

(TABLE 2.3 continued...)

SDS-PAGE 12% Lower gel	10 ml 4x lower buffer 16 ml acrylamide stock 13.8 ml dH_2O 0.2 ml 10% Ammonium persulphate (AMPS)
SDS-PAGE 3% Stacking gel	2.5 ml 4x upper buffer 1 ml acrylamide stock 6.4 ml dH_2O 0.2 ml 10% w/v AMPS
SDS-PAGE 2x Cracking buffer	60mM Tris.Cl (pH 6.8) 1% v/v β-mercaptoethanol 10% v/v glycerol 0.01% v/v bromophenol blue
SDS-PAGE 4x Running buffer stock (RBS)	1.2% w/v Tris base 5.76% w/v glycine
1x Running buffer (RB)	1L 4x RBS 3L dH_2O 40ml 10% w/v SDS

[a] : Constituents in brackets were autoclaved prior to addition and added to other pre-autoclaved constituents.

2.2.2 SUPPLEMENTS TO GROWTH MEDIA

Supplements which were added to liquid media or molten agar media are listed in Tables 2.4 and 2.5 below.

TABLE 2.4 ANTIBIOTIC SUPPLEMENTS FOR CULTURE MEDIA

SUPPLEMENT	STOCK SOLUTION ($mgml^{-1}$) or (w/v)[1]	FINAL CONC. (μgml^{-1})	ABBREVIATION
Sodium ampicillin	5	50	Ap
Kanamycin sulphate	5	50	Kn
Tetracycline[2]	1	10	Tc
Spectinomycin sulphate	5	50	Sp
Chloramphenicol[2]	5	50	Cm

(Table 2.4 continued...)

(Table 2.4 continued...) Rifampicin[2]	20	200	Rf
5-bromo-4-chloro-3-indolyl -D-galacto-pyranoside[3]	30	30	X-Gal
Isopropyl-1-thio-β-D-galactosidase	2.5% (w/v)	As directed	IPTG

Key:

1. All stocks, unless otherwise stated, were filter sterilised
2. Tc, Cm, Rf were made up in 50% v/v ethanol and stored at -20°C in the dark
3. X-Gal was dissolved in dimethylformamide and stored at -20°C in the dark without filter sterilisation.

TABLE 2.5 AMINO ACID SOLUTIONS FOR T7 EXPRESSION STUDIES

AMINO ACID (1% W/V STOCK)	AUTOCLAVED (A) or FILTER STERILISED (F)	NOTES ON PREPARATION
Alanine	A	
Arginine	A	
Asparagine	F	Heat to 60°C to dissolve
[Potassium] Aspartate	F	Heat to 60°C to dissolve
[Sodium] Glutamate	F	
Glutamine	A	
Glycine	A	
Histidine	A	
Isoleucine	A	
Leucine	A	
Lysine	A	
Phenylalanine	A	Dissolve in 0.01M HCl
Proline	A	
Serine	A	
Threonine	A	
Tryptophan	F	
Tyrosine	F	Dissolve with a few drops of 1M NaOH
Valine	A	

2.3 PREPARATION, ANALYSIS AND TREATMENT OF DNA

2.3.1 SMALL-SCALE PREPARATION OF PLASMID DNA (ALKALINE LYSIS MINI PREP)

This protocol is based on the alkaline lysis method as described by Sambrook *et al.* (1989). It was used only where quality of the final preparation was not critical, for example when initially checking DNA from recombinant clones after transformation. The strain was cultured overnight in 10ml LB. 1.5 ml of culture was transferred into a 2 ml Eppendorf tube and centrifuged (13,000 rpm, 2 min, 4°C, MSE Microcentaur). Supernatant was was removed by aspiration to leave a tight pellet, which was resuspended in 100µl of glucose-TE (Table 2.3). To this, 200µl of freshly prepared SDS-NaOH lysis solution was added (Table 2.3), and the tube inverted 5x. 150µl of K-Acetate was added (Table 2.3) to neutralise the lysate and the tube placed on ice for 5 min. The cleared lysate was centrifuged, (13,000 rpm, 3 min, MSE Microcentaur) and the supernatant removed and placed into a fresh Eppendorf tube. An equal of phenol:chloroform (1:1, ~400µl) was added and the tube vortexed thoroughly to create an emulsion. After centrifugation, (13,000 rpm, 3 min, MSE Microcentaur), the organic layer was discarded and 2 volumes of room temperature ethanol was added, followed by vortexing. The tube was then placed at -20°C for 10 min after which it was re-centrifuged (13,000 rpm, 10 min, MSE Microcentaur) and the supernatant removed by aspiration. The pellet was then rinsed in 70% ethanol and recentrifuged as before. Supernatant was removed and the tube left open until the pellet was dry (~10 min). The pellet was then resuspended in sdH_2O and stored at -20°C until use.

Where final preparation quality was required to be high, for example for sequencing, a commercial kit (Qiagen QIAprep Minikit), based on a modified alkaline lysis protocol, but incorporating silica gel anion-exchange resin columns was used according to manufacturer's recommendations. Strains were cultured overnight in LB broth prior to plasmid preparation.

2.3.2 LARGE-SCALE PREPARATION OF PLASMID DNA

The large-scale preparation method was used to obtain large quantities of plasmid DNA for manual sequencing and also for high-yield preparation of cosmids and low-copy plasmids. In such instances, a QIAGEN Midi Kit or a QIAGEN Maxi kit was used according to manufacturer's instruction. Strains were cultured overnight in LB broth prior to plasmid preparation.

2.3.3 PREPARATION OF CHROMOSOMAL DNA

The method is adapted from Sambrook *et al.*, (1989) with some modifications for *S. marcescens* to improve final DNA yield (S. McGowan, Pers. comm.) A 10 ml overnight culture of the strain was cultured and centrifuged at 4,500 rpm, 10 min at 4°C. Cells were resuspended in 10 ml Tris-HCl/EDTA in a glass Universal bottle. 2 ml of lysozyme (Sigma) was added and the tube incubated on ice for 5 min. 200µl of Proteinase K (Boehringer Mannheim) at a concentration of 10mgml-1 and 300µl of 10% SDS were added. The contents were agitated gently to mix and then incubated at 65°C until the cell lysate had cleared. 3 ml of 5M sodium perchlorate was added and the tube returned to 65°C for 1h. The tube was then placed on ice to cool for 10 min. DNA was then extracted once in phenol equilibrated in Tris-HCl (Fisons), then extracted once in 1:1 phenol-chloroform. Finally, DNA was extracted in 24:1 chloroform-isoamyl alcohol. To avoid shearing, the aqueous layer was carefully removed using a trimmed 1 ml Gilson pipette tip and stored at 4°C until use. A 500µl aliquot of the DNA suspension was transferred into an Eppendorf tube and an equal volume of absolute ethanol was added. DNA which precipitated at the interface was condensed into a visible pellet by gently inverting the tube several times. DNA was then centrifuged (MSE Microcentaur, 13,000 rpm, 15 min). After removal of ethanol by crude aspiration, the pellet was left to air dry and resuspended 100µl of sterile water and incubated at 65°C for 15 min. DNA solution was then stored at -20°C until use. This generally yielded enough DNA for 1-2 digestions for *S. marcescens*

2.3.4 PREPARATION OF DNA FOR GENETIC MANIPULATION

2.3.4.1 RESTRICTION ENDONUCLEASE DIGESTION OF DNA

Restriction digestion treatment of DNA (restriction digest) was usually performed in a final volume of 20µl for plasmid DNAs and 30µl for *S. marcescens* chromosomal DNAs. Digestion of DNA was performed according to manufacturer's recommendations (GIBCO BRL, New England Biolabs) in the buffer provided by the manufacturer. Plasmid DNA digestions were done for 1h at 37°C or 2h at 25°C. Chromosomal DNA digestions were done at 37°C for 3h. 'Double digests' using two enzymes were performed simultaneously using New England Biolabs enzymes and the appropriate buffer following manufacturer's recommendations. Where simultaneous digestion was found to be inappropriate, sequential digests were done: DNA was cut with one enzyme under the appropriate conditions, the volume was adjusted to 100µl and the mixture DNA de-salted by using GENECLEAN II™ (BIO101, section 2.3.5.7). DNA was then resuspended in an appropriate volume of sdH_2O and digested with the second enzyme under the appropriate conditions. Efficiency of restriction enzyme digestion was usually verified by agarose gel electrophoresis (section 2.3.5.5).

2.3.4.2 PHOSPHATASE TREATMENT OF DOUBLE-STRANDED DNA

In order to prevent recircularisation of linearised DNA following restriction enzyme digestion, it was dephosphorylated with Calf Intestinal Alkaline Phosphatase (CAP, Boehringer Mannheim). DNA was initially recovered from the digestion mix or from agarose (if extracted as a band following gel electrophoresis) by GENECLEAN II™ (BIO101, section 2.3.5.7), following manufacturer's recommendations. DNA was resuspended in sdH_2O in a final volume of 30µl and digested for 1h at 37°C. The reaction was stopped by repeating the GENECLEAN II procedure and resuspending DNA in an appropriate volume of sdH_2O.

2.3.4.3 FRAGMENT END-REPAIR OF LINEAR DOUBLE STRANDED DNA ("BLUNT-ENDING")

Restriction endonuclease digested DNA that generated 3' or 5' overhangs or sonicated DNA with "ragged ends" was repaired by the use of either T4 DNA Polymerase I (New England Biolabs) or DNA Polymerase I Klenow Fragment (New England Biolabs). Reactions were carried out according to manufacturer's recommendations. dNTP mix (dATP, dCTP, dGTP, dTTP) was obtained by diluting and mixing 100mM stock solutions (Promega) and used at a final concentration of 2.5mM, prediluted in sdH_2O. End-repaired DNA was recovered from the reaction mix by GENECLEAN II ™(BIO101, section 2.3.5.7) according to manufacturer's recommendations and resupended in an appropriate volume of sdH_2O, (usually 14µl for subsequent ligation).

2.3.4.4 LIGATION OF DNA FOR CLONING

The method is essentially that described by Sambrook *et al.* (1989). Vector DNA was digested with the restriction endonuclease(s) of choice, and run out on an electrophoretic agarose mini-gel (section 2.3.5.5). This was done in order to ensure vector DNA had been fully linearised since the presence of non-linearised species would reduce the efficiency of downstream ligation reactions. The linearised vector band was excised from the gel, and extracted by GENECLEAN II™ (BIO101, section 2.3.5.7), according to manufacturer's recommendations, resuspended in 24µl of sdH_2O. It was then dephosphorylated and de-salted as described in section 2.3.5.2. DNA was resuspended in an appropriate volume. "Insert" DNA, to be ligated to the vector, was digested with the required restriction endonuclease(s), run on and excised from an agarose gel (section 2.3.5.5), extracted by GENECLEAN II™ (BIO101, section 2.3.5.7)and resuspended in a minimal volume of sdH_2O, usually 10 -15 µl. 5µl of this was run on an analytical agarose gel (section 2.3.5.8) alongside the same volume of vector DNA in order to estimate relative concentrations of the two species. Vector and insert DNAs were mixed in a 1:3 ratio and an appropriate volume of 10x Ligase Buffer (New England Biolabs). T4 DNA Ligase (New England Biolabs) was added to a final working concentration of 1-2U of enzyme per ligation reaction and sdH_2O was added to bring the final volume to 20µl. Ligation

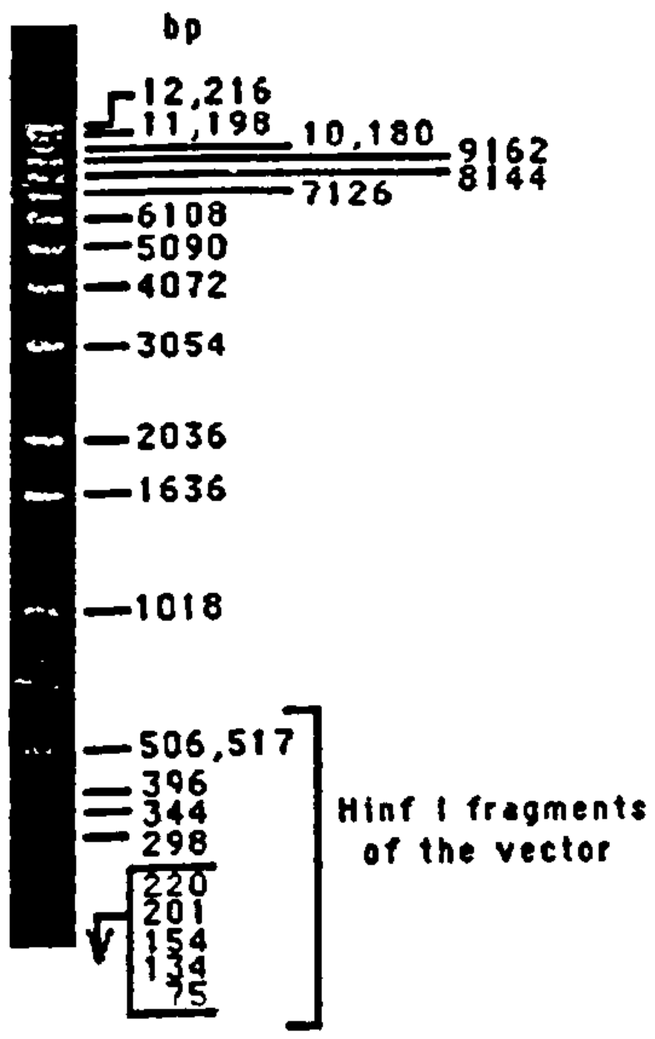
was then performed under recommended conditions, usually overnight at 16°C. ATP required for the reaction was present in the ligation buffer.

2.3.4.5 AGAROSE GEL ELECTROPHORESIS OF DNA

Agarose (GIBCO BRL, Ultrapure) was melted in 1x TBE to a final concentration of 0.7% unless large chromosomal DNA fragments were being cloned, in which case a 1.2% gel was used. Ethidium bromide ($30mgml^{-1}$) was added to the molten agar to a final concentration of $10\mu gml^{-1}$. Molten agar was set in a large gel tray (Biorad) or a minicell with a gel comb. When set, it was submerged in 1x TBE in the gel tank. DNA samples were mixed with 0.2 volumes of Loading Buffer. Gels were run between 10 - 100 V for 1 - 18h as appropriate for gel size and fragment size separation. Sizes of DNA fragments under analysis were determined by comparison with 1Kb Ladder (GIBCO BRL, Section 2.3.4.6). DNA was visualised using a longwave (330nm) UV transilluminator and photographed using Polaroid 665 or 667 film on a shortwave (260nm) UV transilluminator. DNA to be cloned was not exposed to shortwave UV radiation.

2.3.4.6 DNA STANDARDS FOR AGAROSE GEL ELECTROPHORESIS

Size markers for agarose gel electrophoresis of all DNA analysed during this study are shown (manufacturer's information leaflet) below. The 1Kb Ladder (GIBCO BRL) is suitable for sizing double stranded DNA fragments from 500bp to 12Kb, although sizes of fragments <500bp can also be estimated down to approximately 300bp. 5µl of 1Kb ladder was usually run out during electrophoresis, pre-diluted 1:10 in TE buffer to a concentration of 100ngµl^{-1}.



1 Kb DNA Ladder
0.5 µg/lane
0.9% agarose gel
stained with ethidium bromide

2.3.4.7 ISOLATION OF DNA FROM AGAROSE GELS AND RESTRICTION DIGESTION MIXES USING THE GENECLEAN II™ KIT

DNA fragments generated by restriction digest or sonication were separated on an agarose gel. The DNA of interest was visualised on a longwave UV trans-illumminator and excised using a sterilised scalpel blade. DNA was recovered from the gel by GENECLEAN II™ (BIO101) according to manufacturer's instructions. This kit uses sodium iodide and TBE modifier buffer to dissolve the gel slice. DNA is recovered by the addition of a silica gel slurry, incubation, washing and heat treatment.

2.3.4.8 PHENOL-CHLOROFORM EXTRACTION AND ETHANOL PRECIPITATION OF DNA

In most instances where DNA needed to be de-salted in solution, a commercial kit , (GENECLEAN II™, BIO101, section 2.3.4.7), was used (see also Section 2.3). Where this was not used, the method employed is based on that described by Sambrook*et al.* (1989). The DNA solution was combined with an equal volume of Phenol equilibrated with Tris-HCl pH8.0 (Fisons) and vortexed briefly to create an emulsion. The sample was centrifuged at 13,000 rpm and $^3/_4$ of the volume of the upper aqueous layer was carefully removed without disturbing the interface. Small volumes of less than 50µl were diluted by the addition of water to 100µl to ease extraction. An equal volume of chloroform was added and the tube vortexed to create an emulsion. Following centrifugation, the aqueous layer was transferred to a sterile Eppendorf. DNA was recovered by the addition of 0.1 volume of 3M sodium acetate pH 5.5 and two volumes of ethanol, vortexing and chilling at -20°C for 30 min, followed by high speed centrifugation (13,000 rpm, 15 min, MSE Microcentaur). After decanting the supernatant, the pellet was washed in 70% ethanol to remove residual salt and recentrifuged to pellet DNA. The supernatant was discarded and the pellet dried (55°C, 5 min) and resuspended in an appropriate volume of sdH_2O

2.4 TRANSFORMATION OF BACTERIAL CELLS BY EXOGENOUS DNA

2.4.1 CALCIUM CHLORIDE TRANSFORMATION

The method used is essentially that described by Sambrook *et al.* (1989) but modified to further optimise efficiency. The strain was cultured overnight in 2xYT was diluted $^{1}/_{100}$ into 50 ml 2xYT in a 500 ml flask and cultured with vigorous shaking (275 rpm, 37°C, New Brunswick Orbital Shaker or platform shaking water bath). The cells were cultured to OD_{600} =0.9 to improve transformation efficiency, as recommended by Tang *et al.* (1994). The culture was divided into two sterile universal tubes and centrifuged (Labor 50-M, 4,500 rpm, 10 min, 4°C) and the supernatant discarded. Cells were washed in 25 ml of ice cold 0.1M $MgCl_2$ and centrifuged as before. Cells were then washed in 25 ml of ice cold 0.1M $CaCl_2$, centrifuged as before and resuspended in 2.5 ml of ice cold $CaCl_2$. Cell competence was improved by chilling the suspension on ice for at least 3h. Usually, the procedure was carried out the day before the transformation experiment and cells left on ice overnight at 4°C as it has been found that this also improves transformation efficiency (S. Bentley, Pers. comm).

DNA, typically 10-50ng, was added to a 200µl aliquot of competent cells, mixed by flicking the Eppendorf tube and chilled on ice for 40 min-2h. Cells were then heat-shocked at 42°C in a water bath for 2 min and replaced on ice for 3 min. 1 ml of 2xYT was then added and the cells allowed to express for one hour if they were *E. coli* or 2h if they were *S. marcescens* at the appropriate temperature (37°C for *E. coli*, 30°C for *S. marcescens*) prior to spreading on NBA plates with the appropriate antibiotic selection. Competent non-transformed cells from the same stock were used as a negative control.

2.4.2 ELECTROTRANSFORMATION OF BACTERIAL CELLS (ELECTROPORATION)

Electroporation was found to be the most efficient way to transform *E. coli* cells when a high transformation transformation efficiency was critical, for example when using low copy plasmids.

The strain was cultured overnight in 10 ml 2xYT and was diluted $^{1}/_{100}$ in 10 ml of fresh 2xYT in a 100 ml flask and cultured vigorously (250 rpm, New Brunswick Orbital Incubator Shaker, 37°C for *E. coli*, 30°C for *S. marcescens*) to mid-log phase (2-3h for *E. coli*, 4-6h for *S. marcescens*). The culture was centrifuged (Labor-50M, 4,500 rpm, 10 min, 4°C), and washed three times in 10 ml of ice cold sdH_2O. Cells were pelleted as before and resuspended in 1 ml of ice cold sdH_2O and placed on ice. DNA to be transformed into the cells (5-50ng) was pipetted into a 0.2 ml electroporation cuvette (Bio-Rad). To this, 200µl of electrocompetent cells were added, ensuring DNA had been washed down with the introduction of the culture aliquot to the cuvette. The cuvette was flicked to form an even layer at its base and replaced on ice.

Electroporation was carried out using a Bio-Rad Gene Pulser™ according to manufacturer's guidelines. A single 25mF, 2.5kV, 200W pulse was delivered and 1 ml of chilled 2xYT was added immediately after pulsing. Cells were then transferred to a 2 ml Eppendorf and expressed for 1h at 37°C for *E. coli* or 2h at 30°C for *S. marcescens*, and plated on selective medium as appropriate.

2.5 PREPARATION AND USE OF BACTERIOPHAGES

2.5.1 PREPARATION OF HIGH TITRE PHAGE λ LYSATES CARRYING TRANSPOSONS

λTn*phoA*'-2 and λTn*blaM* lysates were prepared on the *E.coli* suppressing strain LE392. An culture of LE392 was cultured overnight in LB broth plus 10mM MgSO4. Bacteriophage carrying the transposon was plaque-purified on a lawn of LE392 as follows: 200µl of the overnight culture of LE392 was mixed with 3ml of molten DDA agar (0.7%) and poured onto a DDA plate and allowed to set. The λ lysate to be purified was serially diluted in phage buffer. 10µl aliquots of neat, 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8} dilutions of lysate were spotted onto the lawn and allowed to dry. The plate was incubated for 16h at 37°C.

After incubation, the dilution yielding discrete single plaques was used. Two plaques (10^{4} - 10^{6} pfu ml^{-1}) were resuspended in 1 ml of phage buffer and the suspension vortexed and left at 4°C for 2h or more. Volumes of 10µl, 50µl and

100µl of phage suspension were added to separate 200µl aliquots of an overnight culture of LE392 and mixed with 3 ml of molten 0.3% DDA. The mixtures were overlayed onto fresh wet chilled DDA plated and incubated at 37°C for 16h.

Following incubation, plates were inspected by comparison to a phage-free control lawn. The plate showing semi-confluent lysis was harvested by removing top agar, washing the plate with 3 ml of phage buffer and pooling in a universal. To this, 0.5 ml of treated chloroform (Table 2.3) was added (pretreated with Na_2CO_3). The lysate was then vortexed for 10 min in order to kill any live cells and then centrifuged to remove cellular debris (Labor-50M, 4,500 rpm, 15min, 4°C). The supernatant was then decanted into a sterile universal and stored over a few drops of chloroform at 4°C until use. The lysate was titrated by spotting 10µl of 10^{-2}, 10^{-4}, 10^{-6} and 10^{-8} dilutions in phage buffer onto a lawn of LE392 cultured in LB + 10mM $MgSO_4$ and incubating overnight at 37°C. Phage lysates were kept on ice at all times during experimental procedures.

2.5.2 TRANSDUCTION OF BACTERIAL CELLS USING PHAGE λ CARRYING A TRANSPOSON (TRANSPOSON MUTAGENESIS)

This method of transduction was communicated by G.P.C. Salmond (unpublished) and has been used successfully for transduction of *E. carotovora* by the generalised transducing phage ΦKP. The *S. marcescens* strain carrying a $LamB^+$ plasmid (pHCP2, pTROY9, pMUT13) was cultured overnight in the presence of antibiotic selection for the plasmid (Ap, Kn or Tc respectively) and 10mM $MgSO_4$. 100µl of culture was removed as a negative control and 100µl of high titre lysate (section 2.5.1) containing $>10^9$ $pfuml^{-1}$ was added. The culture was incubated by shaking at 30°C for one hour, after which it was centrifuged (Labor-50M, 4,500 rpm, 15min) and washed twice in 10 ml of NB, centrifuged, and resuspended in 400µl of NB. Aliquots of 100µl were spread onto NBA plates containing the appropriate selection for the transposon. Unless stated otherwise, plates were incubated at 30°C until discrete colonies appeared (3-4 days for *S. marcescens*).

2.5.3 GENERALISED TRANSDUCTION USING THE *Serratia marcescens* BACTERIOPHAGE ΦOT8

This generalised transducing phage was isolated previously in this laboratory by W. Orme and N. Thomson (unpublished). Preparation of high titre lysates was done as per the method described for phage λ in section 2.5.1, but NBA was used in place of DDA without Mg^{2+} supplement as this is not believed to be required for phage adsorption (G.P.C. Salmond, Pers. comm.). The method of transduction was also the same as that described for phage λ in section 2.5.2.

2.6 THE POLYMERASE CHAIN REACTION (PCR)

The polymerase chain reaction (PCR) was used to amplify a gene(s) of interest from chromosomal DNA samples. The method was essentially that described for the "basic" PCR reaction of Erlich, (1989). Oligonucleotide primers used in PCR experiments are listed in Appendix II. All primers used in this study were specific and non-degenerate. Chromosomal DNA samples were diluted 1:10 in sterile water (Fisons, ultrapure) PCR reagents were added to a 0.5 ml siliconised Eppendorf tube (Sigma), taking care to use a fresh pipette tip for each addition. Reactions were carried out in a total volume of 50μl and overlayed with 50μl of paraffin oil to prevent evaporation during thermal cycling.

Standard PCR reaction mix:-

10 ml each primer
3 ml 100 mM $MgCl_2$ (GIBCO BRL)
2 ml diluted chromosomal DNA[a]
4 ml dNTP mix (dATP, dCTP, dGTP, dTTP; Boehringer Mannheim ,1.25 mM)
5 ml 10x PCR buffer (GIBCO BRL)
16 ml sterile water (Fisons, ultrapure)

Control reactions were also prepared (final volume 50ml made up with sdH_2O) which either lacked template DNA or primers. The PCR machine used was a Omnigene, Hybaid and the standard program used is detailed below:

PCR cycling parameters;

Stage 1

94°C for 5 min - denaturation

Hold at 85 °C - to allow for the addition of 0.3ml Taq DNA polymerase

*? °C for 2 min - Primer annealing temperature

72 °C for 2 min - for extension along the template by Taq polymerase

(1 cycle)

Stage 2

94 °C for 30s - denaturation

*? °C for 30s - annealing

72 °C for 40-60s - to extend, time depending on predicted product size

(35 cycles)

Stage 3

*? °C for 1 min - final annealing

72 °C for 2 min - final extension

(1 cycle)

*? Annealing temperatures for primers were calculated according to the Wallace rule (Ikatura *et al.*, 1984) and are quoted in Appendix II

All products were subsequently run on an agarose gel (section 2.3.5.5).

2.7 SOUTHERN BLOT HYBRIDISATION PROBING

2.7.1 GENERATION OF DIGOXIGENIN-LABELLED PROBE BY RANDOM PRIMED LABELLING

Cosmid DNA from which DNA was obtained for labelling was digested by restriction enzyme digestion (section 2.3.4.1 and analysed by agarose gel electrophoresis (section 2.3.4.5) to ensure it had been digested to completion, and to estimate the amount of DNA present. Non-vector fragments of were excised from the gel and extracted by GENECLEAN II™ (BIO101, section 2.3.4.7). DNA fragments

were labelled with non-radioactive digoxigenin-11-dUTP (DIG) according to manufacturer's instructions, using the random primed DNA labelling kit (Boehringer Mannheim). All probes were stored at -20°C until use and are viable for re-use for up to 12 months.

2.7.2 TRANSFER OF DNA FROM AN AGAROSE GEL TO A NYLON MEMBRANE

The method of transfer of DNA to the filter was carried out essentially as described by Sambrook *et al.*, (1989). Chromosomal DNA to be probed was digested with restriction endonucleases (section 2.3.4.1) and run on an agarose gel (section 2.3.4.5). After electrophoresis, the gel slab was placed in denaturing solution (150 ml in a polypropylene sandwich box) and gently agitated on a platform shaker for 2x 15 min. The DNA was then neutralised by repeating the procedure using Neutralising Solution (150 ml, 2x 15 min). The gel was then placed on filter paper (Whatman 3 mm) on a glass sheet. The filter paper dipped into a tray of containing 200 ml of 10x SSC. Nylon membrane (Hybond-N, Amersham) cut to the size of the gel was placed on top of the gel. Three layers of filter paper cut to size were placed on top of the membrane and soaked in 10x SSC. After ensuring no air bubbles were present between the gel, membrane and filter papers, at least two inches of paper tissue were placed on top and weighted down with a house brick. DNA transfer was left to occur over at least four hours and usually overnight. After transfer, DNA was fixed to the membrane by exposure to short-wave UV radiation (UV transilluminator, 260nm) for 3 min. The membrane was dried and stored between two sheets of Whatman 3 mm paper at room temperature until it was probed.

2.7.3 HYBRIDISATION OF DIG PROBES TO SOUTHERN FILTERS

Hybridisation of Southern blot filters with a digoxigenin-11-dUTP labelled (DIG) DNA probe was carried out essentially according to manufacturers instructions. Solutions used are listed in Table 2.3

Pre-hybridisation and hybridisation were carried out in a hybridisation oven (Hybaid Mini Oven, MKII), according to manufacturers instructions. The hybridisation solution consisted of pre-hybridisation solution plus the DIG labelled

probe. Both hybridisation steps were carried out at a temperature of 65°C unless otherwise stated, and once the DNA probe was added to the filter it was usually left overnight. After the overnight incubation, the probe was recovered and the filter was incubated in Wash I (Table 2.3) at 65°C for 2x 15 min. Subsequent detection steps were carried out at room temperature in a sandwich box. The filter was washed in DIG Buffer I for 1 min and then immersed in 100 ml of DIG Buffer II, with gentle shaking for 30 min. The 'blocked' filter was immersed in 40 ml DIG Buffer II containing 4 ml of anti-DIG antibody for 30 min. The filter was washed twice in DIG Buffer I (2x 100 ml for 15 min each), equilibrated in DIG Buffer III for 2 min and then placed on an acetate sheet (Lloyd Paton). 0.5 ml of DIG Buffer III containing 1:100 dilution of Lumigen PPD detection reagent (kit component) was distributed all over the filter. The filter was then sandwiched between another acetate sheet, heat sealed and exposed to X-ray film (Fuji medical) at room temperature. Exposure time varied between 1-24h.

2.8 SEQUENCING OF DNA

2.8.1.1 PREPARATION OF DNA FOR RANDOM (MANUAL)SEQUENCING

The DNA fragment to be sequenced was obtained by restriction endonuclease(s) digestion (section 2.3.4.1) and isolated from an agarose gel by GENECLEAN II™ (BIO101, section 2.3.4.7). The DNA fragment was self-ligated to generate closed circles due to cohesive ends.

Ligated DNA (in a volume of 30 µl) was sheared using a cup horn sonicator (Heat Systems Ultrasonics W-380) filled with water to a depth of approximately 3 cm, with the sample clamped 1 mm above the probe. DNA was sonicated for two bursts (maximum output, 80s). If necessary between bursts the sample was briefly centrifuged (MSE Micro Centaur, high speed, 3s), replaced on ice, and the water surrounding the probe replaced to aid cooling.

Sheared ends generated by sonication of DNA were repaired to form flush ends (section 2.3.4.2). Repaired sonicated DNA was size-fractionated by gel electrophoresis (section 2.3.4.5) but using high density, 1.5% agarose gel) and DNA

fragments (in the size range 300-1,000bp) were isolated and extracted by GENECLEAN II™ (section 2.3.4.7).

2.8.1.2 LIGATION OF END REPAIRED DNA INTO M13

M13 (50ng, mp18 or mp19, New England Biolabs) DNA was digested with restriction enzyme *Sma*I (GIBCO BRL) at 30°C for 2h, then treated with CAP (section 2.3.4.2). Linearised dephosphorylated M13 was purified from an agarose gel (section 2.3.4.5). End filled, sonicated DNA fragments (section 2.8.1.1) were ligated (section 2.3.4.4) into M13 vector digested with *Sma*I (2.3.41 and 2.3.4.2) and *E.coli* strain TG1, which had been maintained on TG1 maintenance medium agar (Table 2.3), was transformed with the ligated DNA by the $CaCl_2$ method (section 2.4.1). No expression time is required and heat shocked cells were mixed with 3 ml of molten 2xYTA (0.7%) top agar and poured onto freshly prepared 2xYTA plates containing 25µl IPTG and 25µl X-Gal (Table 2.3). Once set, this was incubated overnight at 37°C (blue/white selection, white indicating a disruption to the β-galactosidase gene, therefore likely to be a ligated insert). Competent cells were also transformed with non-digested M13 to gauge transformation efficiency of cells as control. An additional control which was done to assess the efficiency of the cloning procedure was the ligation of *Alu*I-digested λ DNA to prepared M13 vector. This λ DNA ligates with high efficiently into M13 (Bankier Weston and Barrell, Pers. comm.)

2.8.1.3 ISOLATION OF PLAQUES, PREPARATION OF M13 LYSATES AND TEMPLATE DNA

The method used was essentially that of Sanger *et al.* (1977). White plaques were cut from the top agar using a short tipped sterile Pasteur pipette. The plaque was then "blown out" into an Eppendorf containing 50µl of phage buffer (Table 2.3). M13 lysates prepared in this way were viable, when stored at 4°C, for up to two weeks.

An overnight culture of *E.coli* TG1 was diluted $^1/_{100}$ in 2xYT and 1.5 ml aliquots were dispensed into sterile phage tubes. 25µl of the M13 lysate was added to infect the TG1 and incubated at 37°C (New Brunswick G24 Environmental Incubator

Shaker). After a 5h incubation, the culture was centrifuged and the supernatant transferred to another tube containing 200µl PEG precipitation solution (Table 2.3), this was inverted 5x and left overnight at 4°C for the phage to precipitate.

Recombinant phage DNA was recovered by centrifugation (MSE Microcentaur , 13,000 rpm, 20 min) and the supernatant removed by aspiration. A second spin was carried out (MSE Microcentaur , 13,000 rpm, 5 min) to remove any residual PEG precipitation solution and then 100µl of TE was added. The M13 template DNA was then phenol extracted and recovered by ethanol precipitation (section 2.3.4.8) to be resuspended in 25µl of TE and stored at -20°C until use.

2.8.1.4 SEQUENCING RANDOM M13 CLONES USING SEQUENASE™

All sequencing reactions were performed according to manufacturer's instructions (Sequenase™ V2.0 US,Biochemical) using 7µl of the M13 template DNA (section 2.8.1.3). For accuracy and safety, a maximum of 10 templates were sequenced at any one time. Reactions were labelled using ^{35}S-dATP (Amersham). Sequence reactions were stored at -20°C until required and were heated at 80°C for 12 min prior to being loaded onto an acrylamide sequencing gel.

2.8.1.5 CASTING AND RUNNING A POLYACRLYAMIDE SEQUENCING GEL

A Bio-Rad BR102 sequencing rig was prepared and assembled according to manufacturers instructions using 0.4 mm spacers. The acrylamide gel mixes used are as follows:-

1x TBE gel mix (white)	75 ml 40% acrylamide stock(Table 2.3) 50 ml 10x TBE (Table 2.3) 230g urea (BRL) Filtered and stored in the dark at 4°C Made up to 500 ml with water

5x TBE gel mix (blue)	30 ml 40% acrylamide stock 100 ml10x TBE 92 g Urea (BRL) 10 mg bromophenol blue Filtered and stored in the dark at 4°C Made up to 200 ml with water

To cast a gel a plug of acrylamide gel was made between the two gel plates. The plug consisted of 8 ml of 5x TBE mix to which 24 ml of TEMED (Sigma) and 24 ml of ammonium persulphate (AMPS, 20% freshly prepared, Bio-Rad) was added. This was taken up in a 10 ml pipette and run out slowly along the bottom of the glass plates. The gel mix was taken up by capillary action and dried to leave a firm plug in ~20 min. To form the main body of a gradient gel, 75 ml of the 1x TBE (white) and 12 ml of the 5x TBE (blue) acrylamide gel mixes were poured into separate washed beakers. TEMED (150µl) and AMPS (150µl) were added to the white gel mix and 24 ml of TEMED and AMPS were added to the blue gel mix. Reagents were briefly mixed by swirling and then 60 ml of the white mix was taken up in a 100 ml syringe and set aside. 10 ml of the white gel mix was taken up in a 25 ml pipette followed by 12 ml of blue gel mix, two bubbles were drawn into the pipette to mix the two solutions. The acrylamide mix in the 25 ml pipette was run down the centre of the assembled gel rig (held at an angle of ~45°) followed quickly by all the white gel mix in the syringe. A well forming comb was inserted and the gel allowed to set (typically 1h).

The gel plates were inserted into the lower reservoir tray and these reservoirs filled with 1x TBE (table 2.3). The gel comb was removed and the wells flushed with 1x TBE to remove any unpolymerised acrylamide. Sequencing reactions (section 2.8.1.4) were heated at 80°C for 12 min and loaded onto the gel (2µl using a Hamilton syringe). The reactions were separated on the gel for half an hour after the first dye front had run off the bottom of the gel (constant power 100W). The gel was fixed (10% v/v acetic acid 15 min), transferred to Whatman filter paper covered in clingfilm and dried (Bio-Rad Slab Gel Drier, 80°C for 30min). The clingfilm was removed and the gel exposed to X-ray film (Fuji Medical, overnight).

DNA sequence was analysed with the aid of a digitiser (Science Accessories Corporation) and Microgenie computer package (Beckman). The sequence was then analysed on various computer packages which are described where relevant in subsequent chapters.

2.8.2 AUTOMATED SEQUENCING OF DNA BY PRIMER WALKING

The majority of sequencing done in this study was obtained automatically and relied upon the use of specific oligonucleotide primers and plasmid DNA templates.

2.8.2.1 PREPARATION OF PLASMID DNA FOR AUTOMATED SEQUENCING

It was essential to ensure template DNAs were as pure as possible for automated sequencing. DNA was submitted for in-house sequencing in ds plasmid form and was prepared using the Midi Prep kit in sdH_2O (Qiagen, section 2.3.2) to ensure purity. To estimate template DNA concentration, the following method was used (S. Harris, Pers. comm.): A 10µl sample of this was run out by agarose gel electrophoresis against 10µl 1Kb ladder (GIBCO BRL, section 2.3.4.6) (pre-diluted 1:10 in TE buffer to a concentration of $100ng\mu l^{-1}$). DNA concentration of the sample was estimated by visual comparison with the standard in the ladder which represents 1.636Kb. After electrophoresis, template DNA sample volume was adjusted to a concentration of ~100ng μl^{-1} in sdH_2O and 10µl of this was submitted for sequencing.

2.8.2.2 PREPARATION OF OLIGONUCLEOTIDE PRIMERS FOR AUTOMATED SEQUENCING

Oligonucleotide primers were designed as non-degenerate 17mers and were synthesised in-house. Primers were provided in lyophilised form and were reconstituted in 1 ml of sdH_2O. Primers were submitted for sequencing reactions at a 10µM concentration. This was calculated from the molecular weight of each primers by use of the following formula: concentration (mM)= $^{\text{yield x 33}}$ / $_{\text{molecular weight}}$.

Estimate of yield was supplied with the primer following synthesis. All primers used in the sequencing of DNA are listed in Appendix II

2.8.2.3 AUTOMATED SEQUENCING

The sequencing facility used for ds DNA sequencing consisted of a Prism 373a XL Automated DNA Sequencer and a Prism 377 XL Automated DNA sequencer (Applied Biosystems Inc.) Chromatolabelling of reactions was done using ABI Prism Big Dye Terminator kits (Applied Biosystems Inc.) and reactions were carried out on PE/Applied Biosystems 9700 Thermal Cyclers.

2.9 CARBAPENEM DETECTION BIOASSAY

Production of carbapenem by *Serratia* was detected using the *E. coli* strain ESS bioassay, this strain of *E. coli* is super sensitive to β-lactam antibiotics. 200 µl of an overnight culture of the *E. coli* strain ESS was added to 4 ml of molten top agar (NBA) (0.7%) and poured onto a NA plate to form a top lawn. Once set, the strain to be tested was stabbed into, with a sterile tooth pick, or spotted (5 µl) onto this top lawn (spotting culture onto the lawn was found more sensitive). The plates were incubated overnight at 25°C. A clear zone of antibiosis (halo) around the test strain indicated antibiotic production.

2.10 ETHYL METHANE SULPHONATE MUTAGENESIS OF *Serratia marcescens*

The method used was adapted from Forbes and Perombelon (1985). *S. marcescens* strain ATCC39006 was sub-cultured 1:25 from a 10 ml overnight culture in 2xYT into 25 ml of 2xYT in a 250 ml flask. The culture was incubated with shaking until the OD_{600}= 0.6. 500µl of EMS was then added. Cell viability was determined immediately prior to the addition of EMS (t=0 min and every 30 min thereafter), by removal of 100µl samples. Samples were washed twice in EMS Buffer A (Table 2.3) and serial dilutions spread onto NBA plates which were incubated at 30°C for 3 days. Viable counts allowed the plotting of a killing curve, and thus the conditions resulting in 1% survival were determined. The experiment was repeated under 1% survival conditions. Mutagenised cells were washed twice, resuspended in 25 ml of 2xYT and

allowed to recover by further incubation for 1h before serial dilutions were plated out and incubated as required.

2.11 PLOTTING THE GROWTH CURVE OF A CULTURE

The optical density (OD) of a culture was measured at 600 nm on a visible light spectrophotometer (Helios I, Unicam Ltd.), at regular intervals throughout the growth phase, by aseptic removal of 1 ml of the culture into a cuvette (Sarstedt). A graph of log OD_{600} against time could then be plotted. If enzyme assays were also performed, an additional 1 ml sample was removed at each measured time point.

2.12.1 DISRUPTION OF BACTERIAL CELL MEMBRANES USING TOLUENE

This method was used in place of the more conventional one of sonication because it was found to be more efficient and reproducible as a means of disrupting cells of *S. marcescens* prior to performing β-galactosidase assays. This protocol was communicated by N. Thomson (Pers. comm.)

50µl of toluene was added to a 1 ml sample of whole culture in broth in a 1.5 ml Eppendorf tube and vortexed for 30s. The tube was left to stand for 10 min in order to allow toluene to settle out. The tube was then agitated very gently for 2s to dislodge and sunken cells, left for an additional 20s to allow any toluene to settle out again, and the appropriate volume of lysed cell culture was removed to assay for β-galactosidase activity. When assays were performed (section 2.12.2), an absorbance reading at A_{550} was also taken to correct for the presence of cellular debris.

2.12.2 β-GALACTOSIDASE (LacZ) ASSAY

The method, based on that of Miller *et al.*, (1972), was used to assess pigment biosynthetic gene activity in LacZ fusion strains of *S. marcescens*. Samples (1 ml) which had been removed from growth curve analysis (performed as described in section 12.11). β-galactosidase activity. 5-200µl of the toluenised sample (section 2.12.2) was added to Z-buffer (Table 2.3) to give a final volume of 500µl. It was found to be extremely important to mix the constituents of the reaction mix thoroughly. The amount of sample added depended on the activity of that sample.

To measure the β-galactosidase activity accurately, in some cases it was necessary to reduce the amount of sample so that the colour change was not immediate. This was particularly important when a large number of samples were being measured concurrently. A timespan of >30 min to observe a colour change was optimal. All samples were measured in triplicate. The blank consisted of 500 µl of Z-buffer.

Aliquots of ONPG (100 ml; 4 $mgml^{-1}$ in fresh Z-buffer) were dispensed into samples, and the time noted. Samples were incubated (37°C) until a faint yellow colour developed at which point the reactions were stopped by the addition of 250µl of 1M Na_2CO_3, and the time noted. Samples were transferred to cuvettes and analysed on a spectrophotometer (Helios I, Unicam Ltd.) at a wavelength of 420 nm. The β-galactosidase activity was expressed in Miller units ($\Delta A_{420}\, min^{-1}\, ml^{-1}$) correcting for the blank and for cellular debris (A_{550}), and sample volume.

2.13 THE T7 EXPRESSION SYSTEM

Plasmid-encoded proteins were identified by the method of Tabor and Richardson (1985). The gene(s) of interest were cloned into pT7-6 (section 2.3.4) so that expression was under control of the T7 RNA polymerase promoter. The plasmid was transformed by electroporation (section 2.4.2) into *E. coli* K38 carrying pGP1-2 which encodes T7 RNA polymerase, expressed from a promoter only at 42°C due to inactivation of the λC_{I857} repressor. Colonies were cultured on LBA-Kn-Ap at 30°C for 16h.

Single colonies of transformed K38 (pGP1-2, pT76-derivative) were cultured in a 10 ml of LB-Kn-Ap overnight. The following day, overnight culture was diluted 1:40 (0.5 ml in 19.5 ml LB) in a 100 ml flask with Kn and Ap selection and cultured in a shaking waterbath at 30°C, (275 rpm, New Brunswick Aquatherm) until OD at A_{590}= 0.4 with LB as a blank (~2-3h). A 1 ml culture aliquot was then removed and centrifuged (Labor 50-M, 4,500 rpm, 10 min). Cells were then washed with 5 ml of M9 medium (Table 2.3), resuspended in 5 ml M9 medium supplemented with 0.02% amino acids (Table 2.5) <u>without</u> cysteine and methionine, and returned to a 50 ml flask for further incubation (30°C, 275 rpm, 60 min).

Gene expression was induced in the experimental and one of two control flasks by temperature shift (42°C, 275 rpm, 20 min) and then Rf (Table 2.4) was

added to the experimental flasks to a final concentration of 200μgml^{-1}. The culture was left at 42°C for a further 20 min before being returned to 30°C (275 rpm, 20 min).

From each flask, a 0.5 ml sample was transferred to a 2 ml Eppendorf tube. 10μCi of ^{35}S labelled methionine (Redivue™, Amersham) was added and the tube re-incubated at 30°C. Cells from each experiment were centrifuged (MSE Microcentaur, 13K, 1 min) and resuspended in 150μl of cracking buffer (Table 2.3). Samples were stored at -20°C until use, when 40μl aliquots were boiled for 5 min and analysed by SDS-PAGE (section 2.14).

2.14.1 PROTEIN ELECTROPHORESIS (SDS PAGE)

The protocol is based on that of Silhavy *et al.* (1984). The gel apparatus (Bio-Rad) was set up according to manufacturer's instructions with thorough cleaning of gel plate assembly components with ethanol and acetone before use. A 12% acrylamide gel mix (Table 2.3), was prepared and poured between the gel plates leaving ~3 cm at the top for the stacking gel mix. A layer of water-saturated butanol was added with a Pasteur pipette and the gel left to polymerise for 1-2h at room temperature. Butanol was then poured off and the gel rinsed with water. The 3% stacking gel was prepared (Table 2.3) and poured on top and a gel comb (Bio-Rad, pre-wetted with 10% w/v freshly made AMPS) was introduced. The gel was left to polymerise, as before, for 1-2h.

Once set, the gel was placed in a tank of 1x RB, standing on a magnetic stirrer. The comb was carefully removed and the wells rinsed out. Protein samples (40μl) which were pre-boiled (section 2.13) were loaded into the wells. Molecular weight standards (20μl, section 2.14.2) were also loaded and the samples were run into the stacking gel at 40mA. Once the dye front reached the bottom of the stacking gel, current was adjusted to 7mA and electrophoresis continued until the blue dye front reached the bottom of the gel.

After electrophoresis, the gel was removed from the rig and plates and fixed in isopropanol:water:glacial acetic acid (25:65:10) with gentle agitation (Luckham Ltd., platform shaker, half-speed) for 30 min. After transfer to 3mm

Whatman paper, and the gel was vacuum-dried (Bio-Rad, Slab Gel Drier, 65°C, 3h) and exposed to X-ray film (Fuji Medical)

2.14.2 MOLECULAR WEIGHT STANDARDS FOR SDS-PAGE

Markers used were high-molecular weight range (14300-220000) Rainbow™ [14C]methylated protein molecular weight markers (Amersham). The constituents are listed below:

PROTEIN	COLOUR	MW (Da)
Myosin	Blue	220000
Phosphorylase b	Brown	97400
Bovine serum albumin	Red	66000
Ovalbumin	Yellow	46000
Carbonic anhydrase	Orange	30000
Trypsin inhibitor	Green	21500
Lysozyme	Magenta	14300

CHAPTER 3

SEQUENCING AND ANALYSIS OF THE 5' END OF THE PRODIGIOSIN BIOSYNTHETIC CLUSTER

3.1.1 PREFACE

The most extensive experimental undertaking during this study was the sequencing of approximately the first half of the prodigiosin biosynthetic (*pig*) cluster. The aim of this work was to sequence the region of DNA, carried in cosmid pNRT104 (section 1.3.4.2), which carries the structural genes for prodigiosin biosynthesis. Analysis of the DNA sequence of the *pig* cluster allowed identification of putative open reading frames (ORFs), possible promoters and ribosome binding sites (RBSs).

3.1.2 THE PRODIGIOSIN BIOSYNTHETIC CLUSTER

As described previously in Section 1.3.4.2, previous work in this laboratory demonstrated that the genes encoding pigment biosynthesis in *S. marcescens* are located in a cluster with a maximal upper size limit of ~35Kb (Thomson, 1996). Prior to sequencing, the distal ends of the cluster, within the 35Kb of *S. marcescens* DNA carried on pNRT104, were determined by restriction mapping and subcloning (A. Cox, unpublished). This allowed identification of key restriction enzyme sites and defined the size of the *pig* cluster more precisely. Subclones which were generated are represented in Fig. 3.1. All of the subclones were carried on the small, high copy Cm^R vector pDAH330 (Appendix I). This work resulted in the generation of a definitive recombinant plasmid, pM245 (Fig. 3.1), carrying an estimated 25Kb of *S. ma*rcescens DNA. Subclones of pM245 failed to reconstitute pigment production in *Erwinia carotovora* subsp. *carotovora* (A. Cox, pers. comm.).

3.2 SEQUENCING OF THE *pig* CLUSTER

The task of sequencing the *pig* cluster was divided between myself and A. Cox. Based on the locations of *Bam*HI sites in pM245, an arbitrary decision was made for the "first half" of the cluster, from the *Pst*I site (designated nucleotide (nt) 1), to the *Bam*HI site at approximately 11500nt in pM245 (Fig. 3.1) to be sequenced in this study. The latter "half" of *S. marcescens* DNA carried on pM245 was sequenced in parallel by A. Cox.

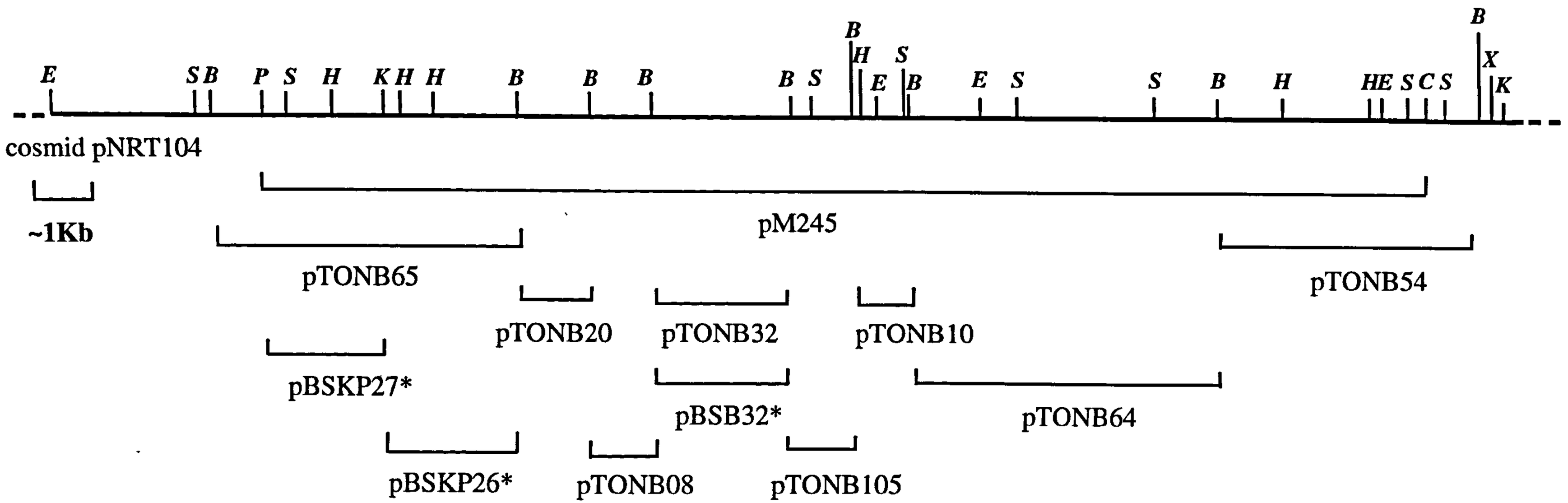


FIGURE 3.1 RESTRICTION MAP AND SUBCLONES OF THE PRODIGIOSIN BIOSYNTHETIC CLUSTER

Pigment production in the *Ecc* heterologous host system was only reconstituted by cosmid pNRT104 and plasmid pM245. Other subclones which were generated are also shown; those marked (*) were generated as part of this study. Key: ***B***: *Bam*HI ***E***: *Eco*RI ***C***: *Sac*I ***H***: *Hin*DIII ***K***: *Kpn*I ***S***: *Sal*I ***X***: *Xba*I

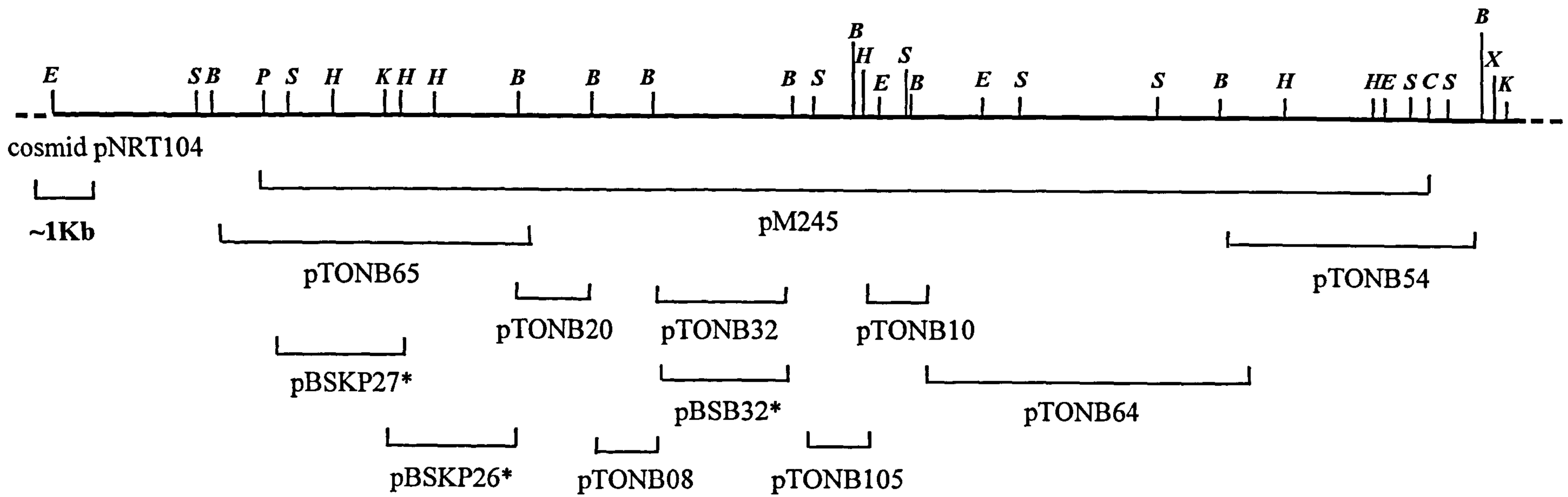


FIGURE 3.1 RESTRICTION MAP AND SUBCLONES OF THE PRODIGIOSIN BIOSYNTHETIC CLUSTER

Pigment production in the *Erwinia carotovora* subsp. *carotovora* heterologous host system was only reconstituted by cosmid pNRT104 and plasmid pM245. Other subclones which were generated are also shown; those marked (*) were generated as part of this study. Key: ***B***: *Bam*HI ***E***: *Eco*RI ***C***: *Sac*I ***H***: *Hin*DIII ***K***: *Kpn*I ***S***: *Sal*I ***X***: *Xba*I

3.2.1 RANDOM CLONING AND MANUAL SEQUENCING

The initial method of choice for sequencing was the M13 dideoxynucleotide chain termination method of Sanger *et al.*, (1977, section 2.8.1). *Bam*HI subclones of the *pig* cluster, in pDAH330 (Figure 3.1), were used to prepare *S. marcescens* DNA for random cloning into M13. As sequencing progressed, a change in experimental approach (Section 3.2.2) meant that the only region of the cluster sequenced in this way during this study, was that carried on the 2.0 Kb subclone pTONB20. *Bam*HI subclone fragments carried in pDAH330 were cloned directly into *Bam*HI digested M13 and sequenced manually, across *Bam*HI junctions from the universal primer sites by A. Cox.

3.2.2 AUTOMATED SEQUENCING BY "PRIMER WALKING"

To hasten sequencing of the *pig* cluster, an automated sequencing facility was used. The *pig* cluster was then sequenced by the method of "primer walking" (Szybalski, 1993; Korneluk *et al.*, 1985). This strategy relies upon the use of oligonucleotide primers designed to the ends of each successive sequence reaction obtained automatically, to sequence further downstream on that strand. Sequences obtained were overlapped and merged using SeqEd software (Applied Biosystems Inc.), which presents automatically generated sequence data in chromatogram format. Finally, University of Wisconsin GCG (UWGCG) software was used to analyse the sequence data.

3.2.2.1 PRIMARY AUTOMATED SEQUENCING REACTIONS: USE OF UNIVERSAL PRIMER SITES

In the initial automated sequence reactions for both strands, a slightly different approach was required, since *S. marcescens* sequence data were not available to design primers for the first round of automated sequencing. The first runs of sequence for both strands were obtained for the ends of the cloned DNA fragments in plasmids pBSKP27, pBSKB26 and pBSB32 respectively (Fig. 3.1). These recombinant plasmids were constructed during this study as follows: DNA from pTONB65 was digested sequentially using *Pst*I and *Kpn*I, and separately,

also digested sequentially using *Kpn*I and *Bam*HI, as described previously (section 2.3.4.1). Digested DNA fragments were separated by agarose gel electrophoresis (section 2.3.4.5) and fragments of ~2.7 Kb (*Pst*I-*Kpn*I) and ~2.6 Kb (*Kpn*I-*Bam*HI) were extracted (section 2.3.4.4). Fragments were ligated to appropriately digested vector, pBluescript™ II KS (Stratagene, Appendix I, section 2.3.4.4), and transformed into *E. coli* DH5α by the $CaCl_2$ method (section 2.4.1). Recombinant clones were selected on NBA-Ap-Xg-IPTG by using blue-white selection. Plasmid pBSB32 was generated by the same method, but the 3.2 Kb insert *S. marcescens* DNA was obtained from *Bam*HI digestion of pNRT104 (Fig. 3.1).

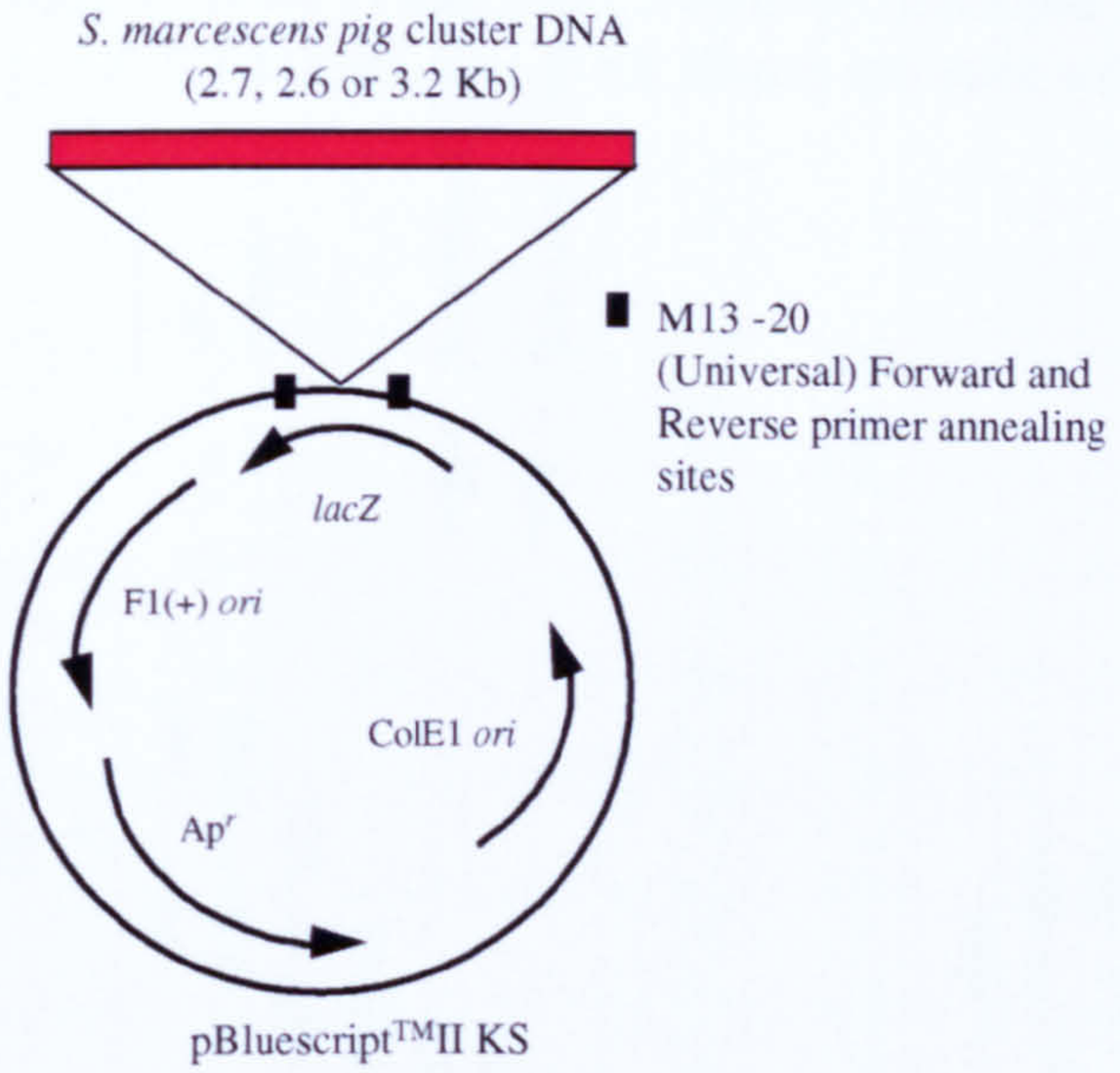
After cloning into pBluescript™ II KS, the "universal" forward (M13 - 20) and reverse primer sites in the *lacZ* gene in the multiple cloning site of the vector were used to obtain primary sequence data for these fragments (Fig. 3.2a). Subsequent sequence reactions were performed on pM245 template, which allowed direct sequencing across the fragment junctions (key restriction enzyme sites) shown in Fig. 3.2b, which greatly aided the compilation of DNA sequence data.

3.2.2.2 THE COMPLETE SEQUENCE OF THE "FIRST HALF" OF THE *pig* CLUSTER

When both strands had been completely sequenced, sequences were merged with those generated by random cloning and sequencing of the 2.0 Kb *Bam*HI fragment, and also with the framework data which were generated by sequencing the ends of *Bam*HI fragments of the *pig* cluster in M13 by A. Cox. Thus, a contiguous sequence was obtained for the first half of the *pig* cluster. A scale representation of the cluster is shown in Fig. 3.3. Following analysis of the sequence (section 3.3), this was designated the "5' end" of the cluster. Fig. 3.3 shows that primer walking was an effective approach. Redundancy of sequence data was necessary because some of the prodigiosin biosynthetic genes might not have database homologues, so it was of paramount importance that the finalised sequence were as accurate as possible. For the region between the first and third *Bam*HI junctions of the *pig* cluster from the left-hand end (Fig. 3.3), the level of redundancy was in fact many-fold greater than depicted in Fig. 3.3,

FIGURE 3.2 CONSTRUCTS AND STRATEGY USED TO GENERATE INITIAL SEQUENCE DATA FOR PRIMER WALKING

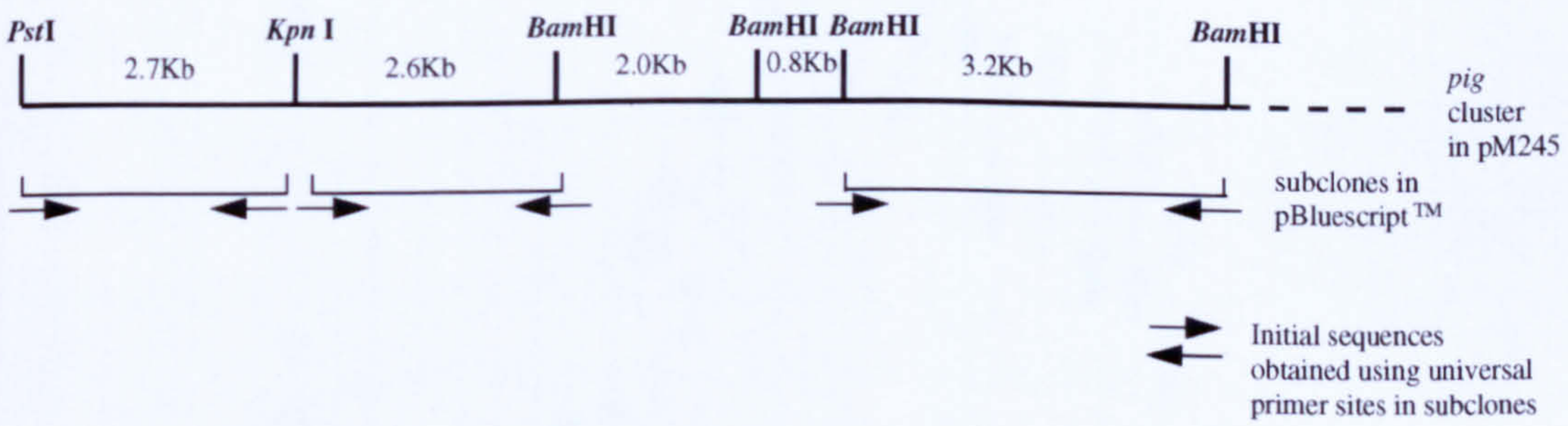
Figure 3.2a



3.2a Plasmids pBSKB26, pBSKP27 and pBSB32

Plasmids were constructed by cloning into pBluescript™II KS. See text in section 3.2.2.1 for details.

Figure 3.2b

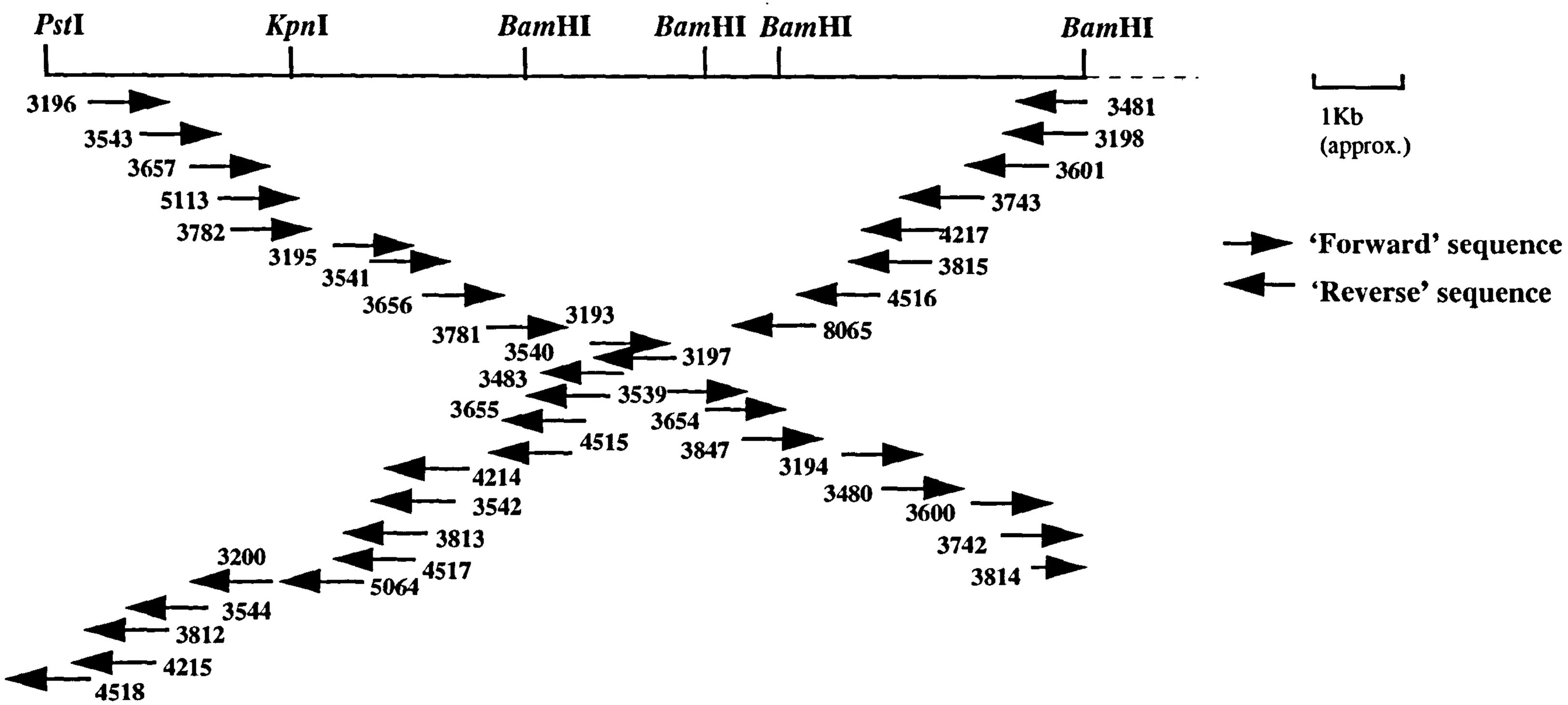


3.2b Restriction Sites and Subclones of the "5' End" of the *pig* Cluster

Automated sequencing began by using the universal forward and reverse primer sites for the first round of sequencing (see Fig. 3.2a). Subsequent primers used for sequencing both strands were designed from data obtained from successive sequence reactions.

FIGURE 3.3 A SCHEMATIC REPRESENTATION OF THE COMPILED SEQUENCE OF THE "5' END" OF THE *pig* CLUSTER IN PLASMID pM245

Numbers associated with arrows refer to 17mer oligonucleotides used to generate each fragment of sequence within the *pig* cluster (listed in Appendix III). Restriction enzyme sites are shown for *S. marcescens* DNA in plasmid pM245.


PstI
KpnI
BamHI
BamHI
BamHI
BamHI
3196
3543
3657
5113
3782
3195
3541
3656
3781
3540
3483
3655
3654
3847
3194
3480
3600
3742
3814
3200
3544
3812
4215
4518
3193
3197
3539
4515
4214
3542
3813
4517
5064
3481
3198
3601
3743
4217
3815
4516
8065
1Kb
(approx.)
'Forward' sequence
'Reverse' sequence

because sequence data obtained from M13 random cloning and sequencing (data not shown) were merged with primer walking sequence data. Finally, the contiguous sequence from this study was merged with sequence data from the latter half of the cluster (A. Cox, Pers. comm.), to obtain a single contiguous sequence for the *pig* cluster.

3.3 ANALYSIS OF THE "5' END" OF THE *pig* CLUSTER

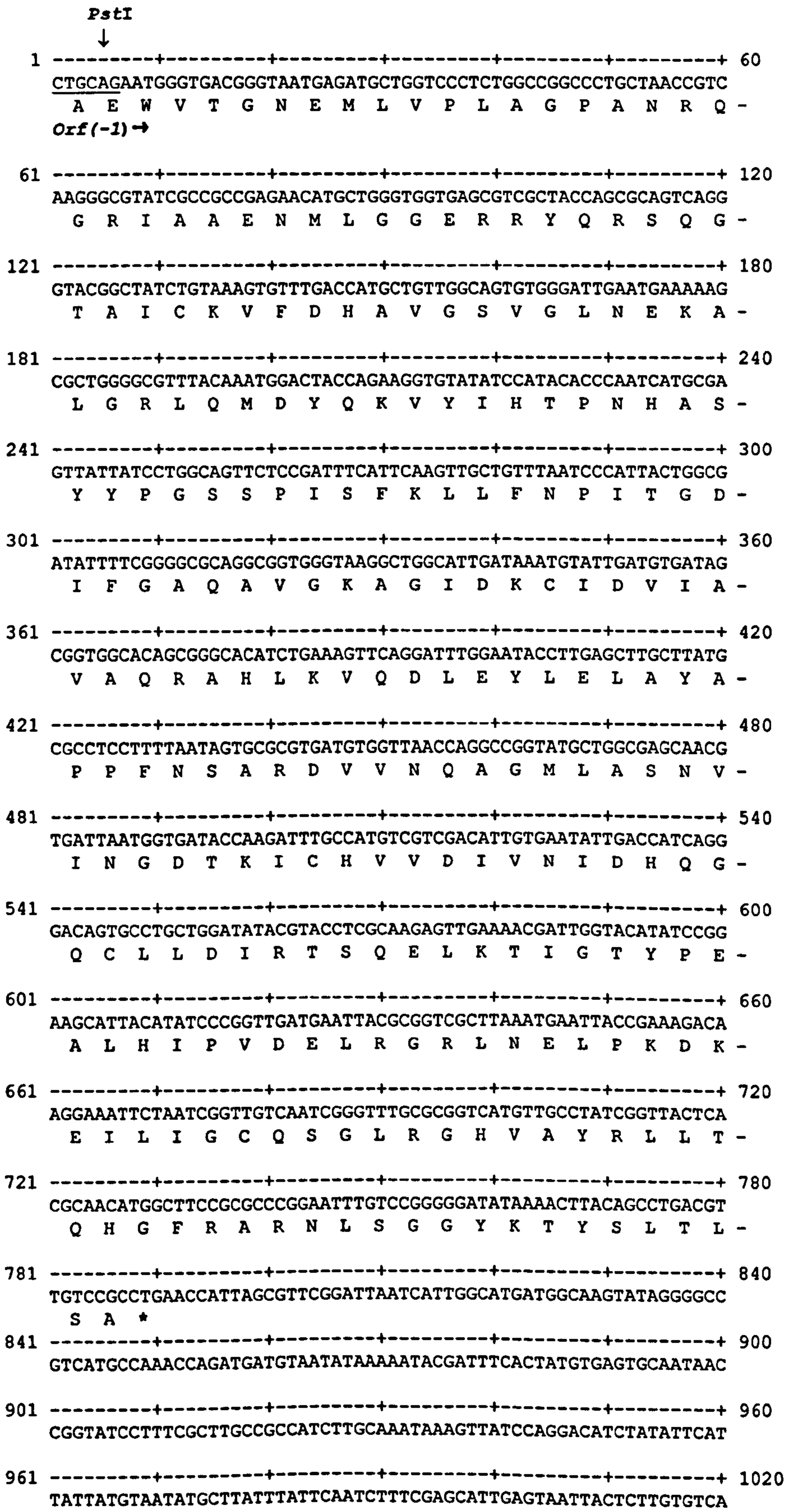
Contiguous nucleotide sequence was analysed using UWGCG software. Six-frame translation of the nucleotide sequence using MAP and PEPDATA analysis to locate any regions bounded by either methionine (ATG) or valine (GTG) initiation codons and stop codons (TAA, TAG, TGA), revealed that likely coding regions existed in all three forward (+) frames. No likely coding regions were present in the opposite orientation and therefore the left-hand end of the *pig* cluster, proceeding from the *Pst*I site in pM245 (Fig. 3.1) to the *Bam*HI site at nt11477 in pM245 was designated the "5' end". The nucleotide sequence (for clarity, non-coding strand only), putative ORFs and possible regulatory motifs (section 3.3.2.) are shown in Fig. 3.4. The putative proteins are discussed in Chapter 4.

Four complete and two partial putative unidirectional ORFs were identified in 11.477 Kb of *S. marcescens* DNA which comprises the 5' end of the *pig* cluster. The first putative ORF (Fig.3.4) is a partial one, of which only the 5' end (in the coding DNA strand) is present in plasmid pM245. It is therefore concluded that Orf(-1) cannot be involved in prodigiosin biosynthesis as pM245 retains the ability to reconstitute pigment production in *Erwinia carotovora* subsp. *carotovora* (A. Cox, pers. comm.). Succeeding *orf(-1)* is an intergenic region of 922nt (Fig. 3.4), which is characterised by a high incidence of stop codons (data not shown). This region was found to contain several interesting and potentially significant features (discussed in section 3.4.4). The first complete ORF of the *pig* cluster, designated *orf1*, starts at nt1713. *orf1* is followed by two more putative ORFs, designated *orf2* and *orf3*, both of which are apparently translationally coupled to *orf1* and *orf2* respectively. This is followed by an intergenic region of 163 nucleotides. This is followed by another putative coding region comprising

FIGURE 3.4 NUCLEOTIDE SEQUENCE AND ORFs OF THE 5' END OF THE *pig* CLUSTER

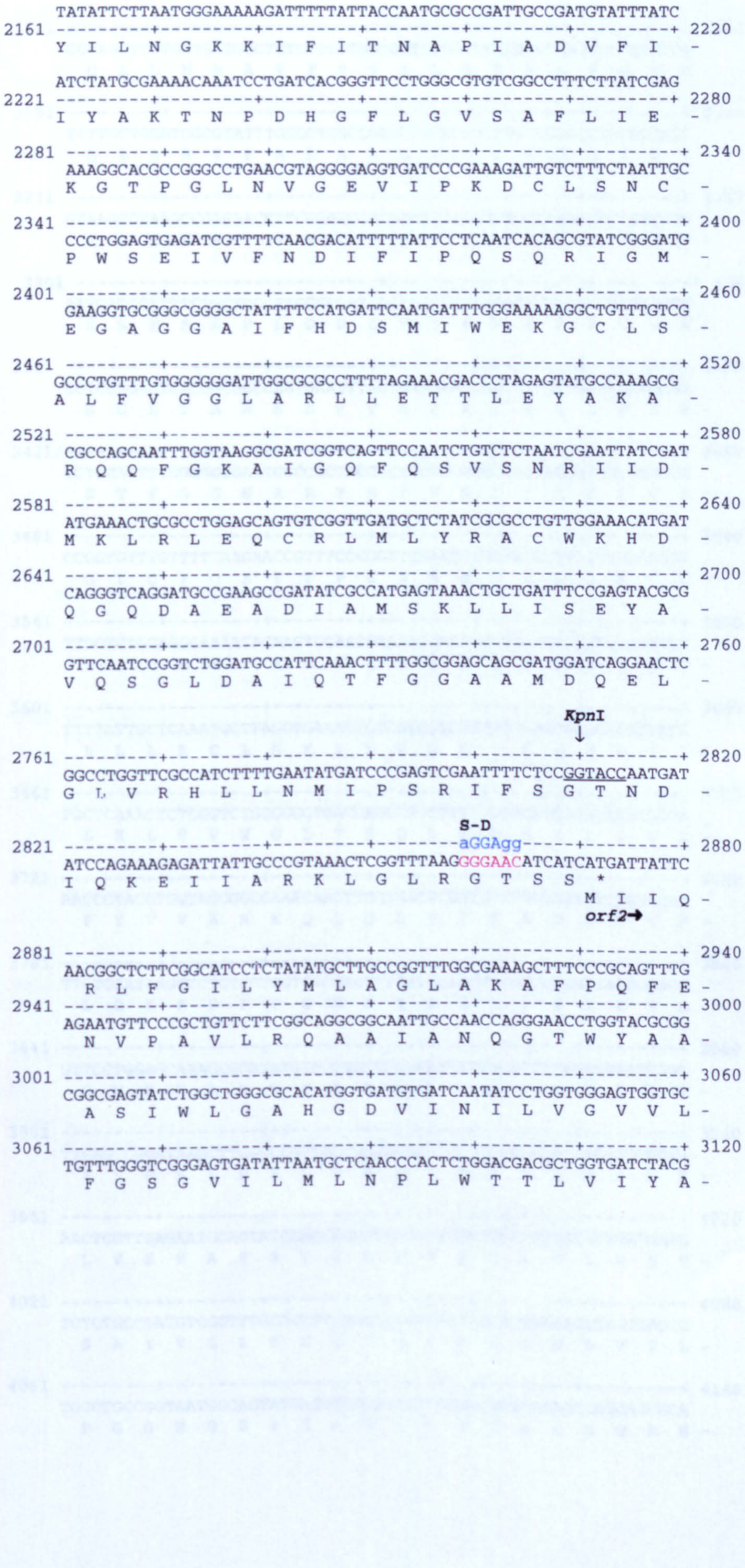
Sequence was obtained up to the *Bam*HI site at 11.478Kb in pM245. A continuation of the contiguous sequence beyond this point is shown for clarity. Putative ribosome binding sites (Shine-Dalgarno sequences, after Shine and Dalgarno, 1974) are highlighted and labelled (**S-D**). Possible σ^{70}-dependent promoter sequences are highlighted, underlined and labelled (**-35**, **-10**). Key restriction enzyme sites, used to aid sequence compilation and for subcloning, are also shown, underlined.

Figure 3.4


PstI
Orf(-1)→
1 CTGCAGAATGGGTGACGGGTAATGAGATGCTGGTCCCTCTGGCCGGCCCTGCTAACCGTC 60
A E W V T G N E M L V P L A G P A N R Q -
61 AAGGGCGTATCGCCGCCGAGAACATGCTGGGTGGTGAGCGTCGCTACCAGCGCAGTCAGG 120
G R I A A E N M L G G E R R Y Q R S Q G -
121 GTACGGCTATCTGTAAAGTGTTTGACCATGCTGTTGGCAGTGTGGGATTGAATGAAAAAG 180
T A I C K V F D H A V G S V G L N E K A -
181 CGCTGGGGCGTTTACAAATGGACTACCAGAAGGTGTATATCCATACACCCAATCATGCGA 240
L G R L Q M D Y Q K V Y I H T P N H A S -
241 GTTATTATCCTGGCAGTTCTCCGATTTCATTCAAGTTGCTGTTTAATCCCATTACTGGCG 300
Y Y P G S S P I S F K L L F N P I T G D -
301 ATATTTTCGGGGCGCAGGCGGTGGGTAAGGCTGGCATTGATAAATGTATTGATGTGATAG 360
I F G A Q A V G K A G I D K C I D V I A -
361 CGGTGGCACAGCGGGCACATCTGAAAGTTCAGGATTTGGAATACCTTGAGCTTGCTTATG 420
V A Q R A H L K V Q D L E Y L E L A Y A -
421 CGCCTCCTTTTAATAGTGCGCGTGATGTGGTTAACCAGGCCGGTATGCTGGCGAGCAACG 480
P P F N S A R D V V N Q A G M L A S N V -
481 TGATTAATGGTGATACCAAGATTTGCCATGTCGTCGACATTGTGAATATTGACCATCAGG 540
I N G D T K I C H V V D I V N I D H Q G -
541 GACAGTGCCTGCTGGATATACGTACCTCGCAAGAGTTGAAAACGATTGGTACATATCCGG 600
Q C L L D I R T S Q E L K T I G T Y P E -
601 AAGCATTACATATCCCGGTTGATGAATTACGCGGTCGCTTAAATGAATTACCGAAAGACA 660
A L H I P V D E L R G R L N E L P K D K -
661 AGGAAATTCTAATCGGTTGTCAATCGGGTTTGCGCGGTCATGTTGCCTATCGGTTACTCA 720
E I L I G C Q S G L R G H V A Y R L L T -
721 CGCAACATGGCTTCCGCGCCCGGAATTTGTCCGGGGGATATAAAACTTACAGCCTGACGT 780
Q H G F R A R N L S G G Y K T Y S L T L -
781 TGTCCGCCTGAACCATTAGCGTTCGGATTAATCATTGGCATGATGGCAAGTATAGGGGCC 840
S A *
841 GTCATGCCAAACCAGATGATGTAATATAAAAATACGATTTCACTATGTGAGTGCAATAAC 900
901 CGGTATCCTTTCGCTTGCCGCCATCTTGCAAATAAAGTTATCCAGGACATCTATATTCAT 960
961 TATTATGTAATATGCTTATTTATTCAATCTTTCGAGCATTGAGTAATTACTCTTGTGTCA 1020

(Figure 3.4 continued...)

(Figure 3.4 continued...)



(Figure 3.4 continued...)

```
3121 ---------+---------+---------+---------+---------+---------+ 3180
     CCCAGCTTTTGATGATGGCTGTCTTTGTGGTGATTTTGCATCAGTCCCAGCCCCAGGTCA
       Q  L  L  M  M  A  V  F  V  V  I  L  H  Q  S  Q  P  Q  V  M -

3181 ---------+---------+---------+---------+---------+---------+ 3240
     TGTTGCTGGATGGCGTATTTGCGCTGGCCGCGCTTTACATGCTGCGCGGTCAGTATCACC
       L  L  D  G  V  F  A  L  A  A  L  Y  M  L  R  G  Q  Y  H  R -

3241 ---------+---------+---------+---------+---------+---------+ 3300
     GTAAGCCTAAGCCGAGAACCTTCCCGACCACGTCTTTTTCGCTGCCCACACCCTCTTCTG
       K  P  K  P  R  T  F  P  T  T  S  F  S  L  P  T  P  S  S  E -

3301 ---------+---------+---------+---------+---------+---------+ 3360
     AATCATCGTTTTCTGCTCCCTTGGGTGATGAGTATGATGTCGTGATTATCGGCGGTGGTG
       S  S  F  S  A  P  L  G  D  E  Y  D  V  V  I  I  G  G  G  A -

3361 ---------+---------+---------+---------+---------+---------+ 3420
     CCTCCGGACTGACGGCGGCCAGTGAATTTACTCATGAACGGGTGTTGGTGCTTGAAAAAA
       S  G  L  T  A  A  S  E  F  T  H  E  R  V  L  V  L  E  K  S -

3421 ---------+---------+---------+---------+---------+---------+ 3480
     GCTCCACTTTCGGGGGGAATGCCCGCTACCACACCTTTAACCGGTTAAAGCATCCCACCG
       S  T  F  G  G  N  A  R  Y  H  T  F  N  R  L  K  H  P  T  A -

3481 ---------+---------+---------+---------+---------+---------+ 3540
     CCGGTGTTTGTTTTCAAGAACCGTTTCCCGGTTCGAATATGCTGCGCTTACTGAAAAAGA
       G  V  C  F  Q  E  P  F  P  G  S  N  M  L  R  L  L  K  K  I -

3541 ---------+---------+---------+---------+---------+---------+ 3600
     TTGGTCTGGAGGGAAAATACAAGTCCAACGAAAAGGACACGCTGGTCTTTTTTGATACCT
       G  L  E  G  K  Y  K  S  N  E  K  D  T  L  V  F  F  D  T  F -

3601 ---------+---------+---------+---------+---------+---------+ 3660
     TTTTATTGCTCAAATGCTTAGGTGAAATTGTGGTCGGCTTTATTAAGCAACCACGTTATC
       L  L  L  K  C  L  G  E  I  V  V  G  F  I  K  Q  P  R  Y  L -

3661 ---------+---------+---------+---------+---------+---------+ 3720
     TGCTCAAACTCTCGGTCTGGGGGCTGACCAGCCAGCTTTTTCTGCATGCGATAATTGGTA
       L  K  L  S  V  W  G  L  T  S  Q  L  F  L  H  A  I  I  G  K -

3721 ---------+---------+---------+---------+---------+---------+ 3780
     AACCCTACGTGGTAGCGGCCAAACAACTTGGTGACCCGATCTTTGCCGATCTTTATACCT
       P  Y  V  V  A  A  K  Q  L  G  D  P  I  F  A  D  L  Y  T  F -

3781 ---------+---------+---------+---------+---------+---------+ 3840
     TTCTGGATAAGTTCTCTCCTCGTGGCGACTTTTATCCACGCCTGCCCTGGACACCGAACG
       L  D  K  F  S  P  R  G  D  F  Y  P  R  L  P  W  T  P  N  G -

3841 ---------+---------+---------+---------+---------+---------+ 3900
     GTTCCTGGAGCAAAGCGCATATGGAGTTGCTCGATAATATTTCCCTTTACACCTATTTGT
       S  W  S  K  A  H  M  E  L  L  D  N  I  S  L  Y  T  Y  L  F -

3901 ---------+---------+---------+---------+---------+---------+ 3960
     TTGAGCCGGATAAGCTTGGCCGGTTACCGGAACAGCTGCGGCCACCCGCCAGACTGGGTA
       E  P  D  K  L  G  R  L  P  E  Q  L  R  P  P  A  R  L  G  K -

3961 ---------+---------+---------+---------+---------+---------+ 4020
     AACTGGTTGAGAATGCAGTATCCACCACCTTGCGGGTTGAGTGTCTGGATATTCATGATG
       L  V  E  N  A  V  S  T  T  L  R  V  E  C  L  D  I  H  D  V -

4021 ---------+---------+---------+---------+---------+---------+ 4080
     TCTCTGCCTATGTCGGTTTGCACTTTCTGGTGGGCTATCTGCGCGGAAACCTGGTCACCC
       S  A  Y  V  G  L  H  F  L  V  G  Y  L  R  G  N  L  V  T  L -

4081 ---------+---------+---------+---------+---------+---------+ 4140
     TGCCTGGCGGTAATGGCAGTATCAGTGCGGGGTTGTGTAAGTATCTAAGCCATCAGCGCA
       P  G  G  N  G  S  I  S  A  G  L  C  K  Y  L  S  H  Q  R  N -
```

(Figure 3.4 continued...)

```
4141 --------+---------+---------+---------+---------+---------+ 4200
     ATGTCACGTTGCAAAACCATGTCCAGTTGACAGCGGTTGAGCCGCAGCACAATGGTACAT
       V   T   L   Q   N   H   V   Q   L   T   A   V   E   P   Q   H   N   G   T   C -

4201 --------+---------+---------+---------+---------+---------+ 4260
     GCATCCAGTTCACGATCAATGGTCAACCCCGCCAAGTGCAGGCTCAACAGATCATTTGGG
       I   Q   F   T   I   N   G   Q   P   R   Q   V   Q   A   Q   Q   I   I   W   A -

4261 --------+---------+---------+---------+---------+---------+ 4320
     CCGCGCCTAAAACGCAACTTGCCACATGGCTGCCGGGATTGCCGGCCAAACAGTTGGCGG
       A   P   K   T   Q   L   A   T   W   L   P   G   L   P   A   K   Q   L   A   A -

4321 --------+---------+---------+---------+---------+---------+ 4380
     CCATCAAGAACATTCGTCATGAAGACTACTATCTGGCGAATGTATTCCTGTCAAAACCGG
       I   K   N   I   R   H   E   D   Y   Y   L   A   N   V   F   L   S   K   P   V -

4381 --------+---------+---------+---------+---------+---------+ 4440
     TGCTGGGTCATTCGTTTGGTGGCTATATGATCGAACCGGACAGCAATAAAGATCCGTTCT
       L   G   H   S   F   G   G   Y   M   I   E   P   D   S   N   K   D   P   F   S -

4441 --------+---------+---------+---------+---------+---------+ 4500
     CTTGGTGTAAAGCGGGGACTTGCCTGGTGGCCAACTGGATGGACGATCATGCTGACGTGG
       W   C   K   A   G   T   C   L   V   A   N   W   M   D   D   H   A   D   V   D -

4501 --------+---------+---------+---------+---------+---------+ 4560
     ATGTGGGCGTGCTGACATTGCTTAAACCCACGACGCGCTCAGAGCGGCAAGACCGTACCG
       V   G   V   L   T   L   L   K   P   T   T   R   S   E   R   Q   D   R   T   A -

4561 --------+---------+---------+---------+---------+---------+ 4620
     CACAGAATGCGTTTTTAGCGCTACAACAGCAAACCTATGCCGAAATTGCCAAGGTTCTGC
       Q   N   A   F   L   A   L   Q   Q   Q   T   Y   A   E   I   A   K   V   L   R -

4621 --------+---------+---------+---------+---------+---------+ 4680
     GCAACATAGGGATTGGTGCCGAGGTTATTGAGGATATCCAGATCTGGTATTGGCCCGCAG
       N   I   G   I   G   A   E   V   I   E   D   I   Q   I   W   Y   W   P   A   G -

4681 --------+---------+---------+---------+---------+---------+ 4740
     GGCTGGTGACGTCGGTAGTGGGGCAACAGGCTGAAGGTGTATTCGAAACCGCCCGTCAGT
       L   V   T   S   V   V   G   Q   Q   A   E   G   V   F   E   T   A   R   Q   S -

4741 --------+---------+---------+---------+---------+---------+ 4800
     CATTTGAAAATATTCATTTCGCTAATCAGGACAGTGTCGGTGTCGGCAATATCGAGAGTG
       F   E   N   I   H   F   A   N   Q   D   S   V   G   V   G   N   I   E   S   A -

4801 --------+---------+---------+---------+---------+---------+ 4860
     CCATCCTGTCGGGAATTGATGCAGCCAACGCGGTTAAAGCGCAACTCATGGATACGGAAA
       I   L   S   G   I   D   A   A   N   A   V   K   A   Q   L   M   D   T   E   N -

           S-D
4861 ------aGGAGG-------+---------+---------+---------+---------+ 4920
     ATGTCGTGGAGGTGGCGGGATGAATCAACCACTGGTTGTTGAAATATCCGGTGATAAGGC
       V   V   E   V   A   G   *
                           M   N   Q   P   L   V   V   E   I   S   G   D   K   A -
                       orf3→

4921 --------+---------+---------+---------+---------+---------+ 4980
     GCTTGAACATCACCACTTGGGCGGTAAGGGTTATTCACTCAACAACCTGATTCATGCGGG
       L   E   H   H   H   L   G   G   K   G   Y   S   L   N   N   L   I   H   A   G -

4981 --------+---------+---------+---------+---------+---------+ 5040
     CCTGCCAGTACCTTCGGCATTTTGTGTGACGGCACAAGCCTACCAACAGTTTATTGAAGA
       L   P   V   P   S   A   F   C   V   T   A   Q   A   Y   Q   Q   F   I   E   E -

5041 --------+---------+---------+---------+---------+---------+ 5100
     GGTGGTGCCGGGGGCGGAATTAACTGACGGCGACCTGATTGCGGTGCGTGACGCCATCCT
       V   V   P   G   A   E   L   T   D   G   D   L   I   A   V   R   D   A   I   L -
```

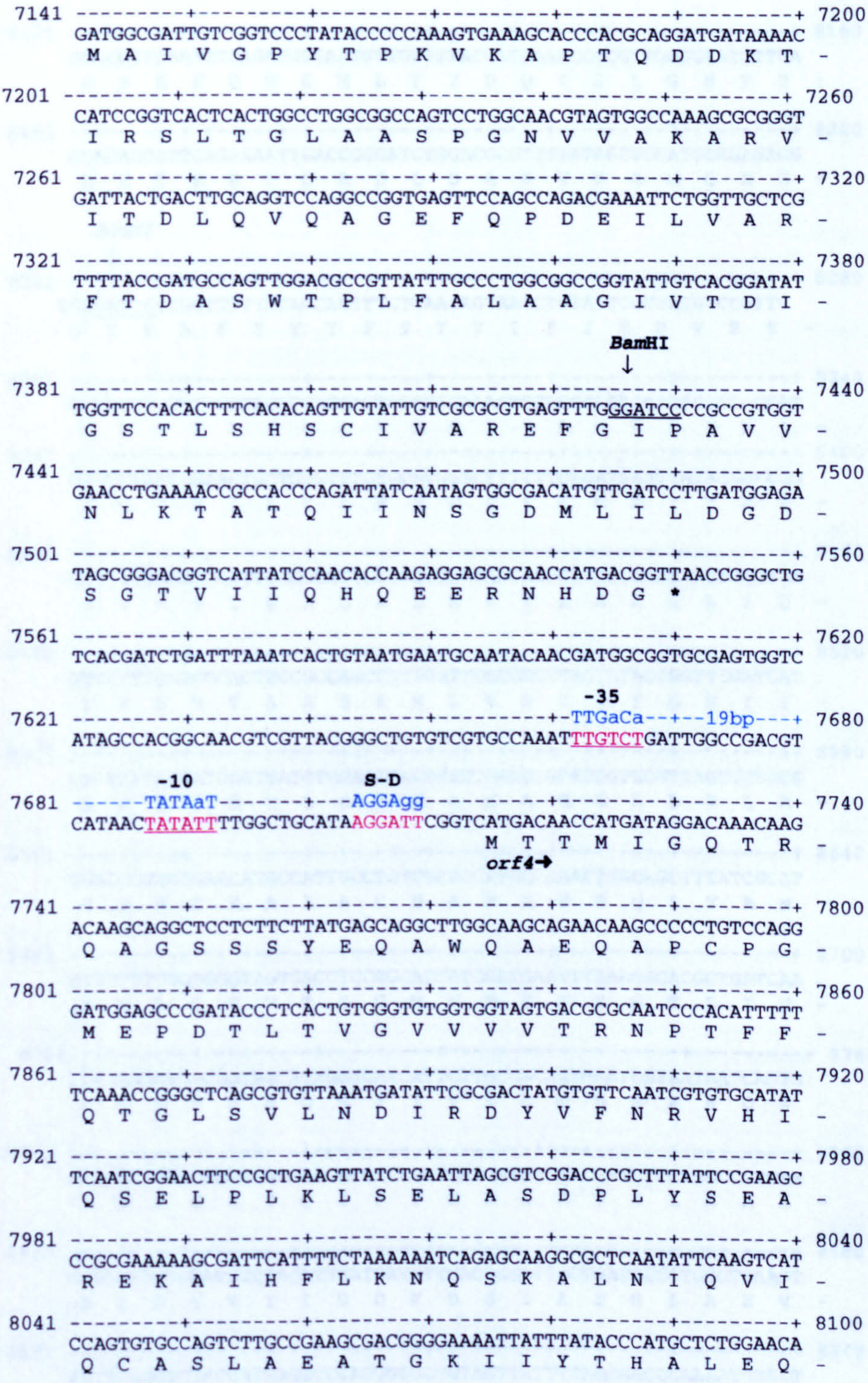
(Figure 3.4 continued...)

```
5101 --------+--------+--------+--------+--------+--------+ 5160
     GCATGCCGATATTCCCGACTCGCTAAAGCAGGCTATCGGTGATGCTTATCAACACCTGGG
      H  A  D  I  P  D  S  L  K  Q  A  I  G  D  A  Y  Q  H  L  G  -

5161 --------+--------+--------+--------+--------+--------+ 5220
     CCATGACACGACCATTGCGGTTCGTTCTTCGGCATTGGATGAAGATGGTCAACGTCAGTC
      H  D  T  T  I  A  V  R  S  S  A  L  D  E  D  G  Q  R  Q  S  -

5221 --------+--------+--------+--------+--------+--------+ 5280
     ATTTGCCGGGCAGTACGAGACTTACCTGCATGTGAAGGGTTCAGAGGCAGTGTTGCACAA
      F  A  G  Q  Y  E  T  Y  L  H  V  K  G  S  E  A  V  L  H  K  -

5281 --------+--------+--------+--------+--------+--------+ 5340
     GGTTCAGGCTTGTTGGGCGTCGCTTTGGGCTGAGCGAGCGGCTCAATATCGTCATGAATC
      V  Q  A  C  W  A  S  L  W  A  E  R  A  A  Q  Y  R  H  E  S  -

5341 --------+--------+--------+--------+--------+--------+ 5400
     TGCGTCGCACAGTGCCATTGCCGTGATTCTACAAGTGATGGTTGATGCCGATGCAGCGGG
      A  S  H  S  A  I  A  V  I  L  Q  V  M  V  D  A  D  A  A  G  -

                    BamHI
                      ↓
5401 --------+--------+--------+--------+--------+--------+ 5460
     AGTGATGTTTACTCAGGATCCTCTGTCAGGGAGCACCGATAAGGTGGTGATTGACAGTTG
      V  M  F  T  Q  D  P  L  S  G  S  T  D  K  V  V  I  D  S  C  -

5461 --------+--------+--------+--------+--------+--------+ 5520
     TTGGGGGCTGGGGGAAGGCGTGGTTTCCGGGCAAGTCACCACAGATAGTTTCACGCTGGA
      W  G  L  G  E  G  V  V  S  G  Q  V  T  T  D  S  F  T  L  D  -

5521 --------+--------+--------+--------+--------+--------+ 5580
     TAAAGCCACCGGTGAGCTATGTGATCAGCAGATTCGCCACAAACCGAATTACTGCCAGCG
      K  A  T  G  E  L  C  D  Q  Q  I  R  H  K  P  N  Y  C  Q  R  -

5581 --------+--------+--------+--------+--------+--------+ 5640
     GGACGAACATGGTCTGGTGACACTGTTGCAAACCCCGGAGGCCAAACGGGATCTCCCCAG
      D  E  H  G  L  V  T  L  L  Q  T  P  E  A  K  R  D  L  P  S  -

5641 --------+--------+--------+--------+--------+--------+ 5700
     TCTGACGCCGGCTCAGTTGCAACAGTTGGTCACGCTGGCCAGACAGGCCCAGCTTATCTA
      L  T  P  A  Q  L  Q  Q  L  V  T  L  A  R  Q  A  Q  L  I  Y  -

5701 --------+--------+--------+--------+--------+--------+ 5760
     CAGCACCGAACTGGATATTGAATGGGCGGTGAAAGATGACAAAGTCTGGTTATTACAGGC
      S  T  E  L  D  I  E  W  A  V  K  D  D  K  V  W  L  L  Q  A  -

5761 --------+--------+--------+--------+--------+--------+ 5820
     GCGTCCGGTAACCACATCGGCCAAAACGGCTAACGTTATCTACGCCAATCCGTGGGAGAG
      R  P  V  T  T  S  A  K  T  A  N  V  I  Y  A  N  P  W  E  S  -

5821 --------+--------+--------+--------+--------+--------+ 5880
     CGACCCGGCTGCGAAGGAGGGCGCTTTTTTCTCGCGGATGGATACCGGAGAGATTGTGAC
      D  P  A  A  K  E  G  A  F  F  S  R  M  D  T  G  E  I  V  T  -

5881 --------+--------+--------+--------+--------+--------+ 5940
     GGGGCTGATGACGCCACTGGGGTTGTCATTTTGTCAGTTCTATCAAAAGCACATTCATGG
      G  L  M  T  P  L  G  L  S  F  C  Q  F  Y  Q  K  H  I  H  G  -

5941--------+--------+--------+--------+--------+--------+ 6000
     CCCGGCCATCAAGACCATGGGGCTTGCCGATATTAGTCACTGGCAGATCTATATGGGCTA
      P  A  I  K  T  M  G  L  A  D  I  S  H  W  Q  I  Y  M  G  Y  -

6001 --------+--------+--------+--------+--------+--------+ 6060
    TATCCAGGGCTATGTTTATCTGAATATTTCCGGTTCGGCCTATATGCTCAGGCAGTGTCC
     I  Q  G  Y  V  Y  L  N  I  S  G  S  A  Y  M  L  R  Q  C  P  -

6061 --------+--------+--------+--------+--------+--------+ 6120
     ACCGACCCGTAATGAAATGAAGTTCACCACCCGCTATGCCACGGACGAGATCGATTTTAA
      P  T  R  N  E  M  K  F  T  T  R  Y  A  T  D  E  I  D  F  K  -
```

(Figure 3.4 continued...)

```
6121 ---------+---------+---------+---------+---------+---------+ 6180
     GGACTATAAAAACCCCTATGGTGCAGGCGTGCAGGGCTGGGATTATGCCAAGAGTTGTTG
      D  Y  K  N  P  Y  G  A  G  V  Q  G  W  D  Y  A  K  S  C  W  -

6181 ---------+---------+---------+---------+---------+---------+ 6240
     GTACTGGCTAAAACAGCAAGTCCGTAATATGCGCAGCGCCGCCAGGACCGTCGAGCAGAT
      Y  W  L  K  Q  Q  V  R  N  M  R  S  A  A  R  T  V  E  Q  M  -

6241 ---------+---------+---------+---------+---------+---------+ 6300
     GATTGCCCTGCGTCAGGACGAAACCACACGGTTTCTGGGGCTTGATTTGACCGCCATGAC
      I  A  L  R  Q  D  E  T  T  R  F  L  G  L  D  L  T  A  M  T  -

6301 ---------+---------+---------+---------+---------+---------+ 6360
     GCTTCAGCAGTTGGACCAAGAGCTGCAACGGATTGACCGCTTTTTCCTCGATAGCTGTGC
      L  Q  Q  L  D  Q  E  L  Q  R  I  D  R  F  F  L  D  S  C  A  -

6361 ---------+---------+---------+---------+---------+---------+ 6420
     GGCGTATATGCCTTTTTTCCTGCAATCATTTGCCCTTTACGATGCGCTGGCGCAAGCCTG
      A  Y  M  P  F  F  L  Q  S  F  A  L  Y  D  A  L  A  Q  A  C  -

6421 ---------+---------+---------+---------+---------+---------+ 6480
     TGAACGCCATATCAAGGATGGTAAAGGTCTGCAAAATCGCATCAAGGCGTCAATGAATAA
      E  R  H  I  K  D  G  K  G  L  Q  N  R  I  K  A  S  M  N  N  -

6481 ---------+---------+---------+---------+---------+---------+ 6540
     CCTGCGCACGATCGAAGTGACGCTCGGTATCATCAAATTGGTTGCGACAGTTAACCAACA
      L  R  T  I  E  V  T  L  G  I  I  K  L  V  A  T  V  N  Q  Q  -

6541 ---------+---------+---------+---------+---------+---------+ 6600
     GACTGAATTAAAAGCGTTATTTGAACAGCATCGTGCCGATGAGTTAGTCACCTTGCTACC
      T  E  L  K  A  L  F  E  Q  H  R  A  D  E  L  V  T  L  L  P  -

6601 ---------+---------+---------+---------+---------+---------+ 6660
     TGTCCATGACATATCCCGTGCTTTCTGGCAGGGTGATTTTGAAGATTTTCTGGTTGAATT
      V  H  D  I  S  R  A  F  W  Q  G  D  F  E  D  F  L  V  E  F  -

6661 ---------+---------+---------+---------+---------+---------+ 6720
     TGGTTCTCGCGGTCGCCAGGAGTTTGATCTGAGTATCCCGCGTTGGCGTGATGACCCGAG
      G  S  R  G  R  Q  E  F  D  L  S  I  P  R  W  R  D  D  P  S  -

6721 ---------+---------+---------+---------+---------+---------+ 6780
     TTATTTATTGCAAGTAATGAAGATGTATTTGCAACACCCGGTAGATTTGCACAAGAAGCT
      Y  L  L  Q  V  M  K  M  Y  L  Q  H  P  V  D  L  H  K  K  L  -

6781 ---------+---------+---------+---------+---------+---------+ 6840
     CAGGGAAACGGAACTGCTGCGTCAGCAGGATAGTGAAGCACTGTTCAGTGCGATGTCCTG
      R  E  T  E  L  L  R  Q  Q  D  S  E  A  L  F  S  A  M  S  W  -

     GTCGGGCCGTTTCAAGCTCAAAACCCTCATCAAGCTGTATGGCATGATGGCCGAACGTCG
6841 ---------+---------+---------+---------+---------+---------+ 6900
     S  G  R  F  K  L  K  T  L  I  K  L  Y  G  M  M  A  E  R  R  -

6901 ---------+---------+---------+---------+---------+---------+ 6960
     TGAAGCGACGCGGCCAACGTTTATCACCGAAACCTGGTTCTACCGTTGCATCATGCTGGA
      E  A  T  R  P  T  F  I  T  E  T  W  F  Y  R  C  I  M  L  E  -

6961 ---------+---------+---------+---------+---------+---------+ 7020
     AGTTCTACGGCGTCTGGACGCGCAGGGGATTGCCAGTAGTGCCGATCTGCCCTATGTCGA
      V  L  R  R  L  D  A  Q  G  I  A  S  S  A  D  L  P  Y  V  D  -

7021 ---------+---------+---------+---------+---------+---------+ 7080
     TTTTGAACAGTTCCGTGCGTATGTGGCGGGGACCATCCCAGCTGAACAGGCGTTTTCCAA
      F  E  Q  F  R  A  Y  V  A  G  T  I  P  A  E  Q  A  F  S  K  -

7081 ---------+---------+---------+---------+---------+---------+ 7140
     AGCGCGGCTCGATCAGAATCGCCATCAACATTTATTCAATTTACATGCCGAAGAACCGCC
      A  R  L  D  Q  N  R  H  Q  H  L  F  N  L  H  A  E  E  P  P  -
```

(Figure 3.4 continued...)



(Figure 3.4 continued...)

```
8101 ---------+---------+---------+---------+---------+---------+ 8160
     GCAGCCTGAATTTCAGATGGGAATGTTGTTTTACGATCAAACCTCGTTGGGGAATGTTGA
      Q  P  E  F  Q  M  G  M  L  F  Y  D  Q  T  S  L  G  N  V  D  -

8161 ---------+---------+---------+---------+---------+---------+ 8220
     CGACAGCATTGAGAAAATTGACCGGGATCTGGACGCGTTTTATAGCGCCATGCAGAGAGG
      D  S  I  E  K  I  D  R  D  L  D  A  F  Y  S  A  M  Q  R  G  -

        BamHI
          ↓
8221 ---------+---------+---------+---------+---------+---------+ 8280
     TGGGATCCCCGCTTTTTATACCACTTTCTCAACAGTAACCTTTATTCGTGATGTCCGTTC
      G  I  P  A  F  Y  T  T  F  S  T  V  T  F  I  R  D  V  R  S  -

8281 ---------+---------+---------+---------+---------+---------+ 8340
     GTCGTTTCGTTATCTACCCCAGCAGTATCGAGAAATTGTCCGCAGTGAAGACCCCGCCAT
      S  F  R  Y  L  P  Q  Q  Y  R  E  I  V  R  S  E  D  P  A  I  -
8341 ---------+---------+---------+---------+---------+---------+ 8400
     CTTCCAAACCGAACTGCTGTGTTTGTGGATGGACTTTTTTGAGATGAATTATACCAACCG
      F  Q  T  E  L  L  C  L  W  M  D  F  F  E  M  N  Y  T  N  R  -

8401 ---------+---------+---------+---------+---------+---------+ 8460
     CCGTGTCAAACCGATTGGGGCGTTGGCACTTCATAACACCCTGGCCGAGCAACTGATTCA
      R  V  K  P  I  G  A  L  A  L  H  N  T  L  A  E  Q  L  I  Q  -

8461 ---------+---------+---------+---------+---------+---------+ 8520
     GTTCTTTGAGCGCACTGCCGCCAGCCGTTGGTTGGTGTCTTACTATACCGGTTCGATCAT
      F  F  E  R  T  A  A  S  R  W  L  V  S  Y  Y  T  G  S  I  I  -

8521 ---------+---------+---------+---------+---------+---------+ 8580
     CTCTAATCTGATCGGTTATCTGGACAGACATGCTGAAGCGCATGGTGCGTTAGTTCTGCG
      S  N  L  I  G  Y  L  D  R  H  A  E  A  H  G  A  L  V  L  R  -

8581 ---------+---------+---------+---------+---------+---------+ 8640
     TGGCCCTAACGAACATGCCATTGCCTGTGGCGCCATGGCGAACTGGCAGCTTTATCGCAT
      G  P  N  E  H  A  I  A  C  G  A  M  A  N  W  Q  L  Y  R  M  -

8641 ---------+---------+---------+---------+---------+---------+ 8700
     GCCTTTTCTGGGGGTAGTGACCTCCGGCATGATGGATGAATTTAAAGGCACGCTGATCAA
      P  F  L  G  V  V  T  S  G  M  M  D  E  F  K  G  T  L  I  N  -

8701 ---------+---------+---------+---------+---------+---------+ 8760
     TCTGAAAGAGACGGCAGCACAGGGCATCATTGTTGCCGCAGAGAATCGTAATAATCAGTG
      L  K  E  T  A  A  Q  G  I  I  V  A  A  E  N  R  N  N  Q  W  -

8761 ---------+---------+---------+---------+---------+---------+ 8820
     GTATAGCTTCCAGGGCACACAGACGCCGACTGAAGATATGCGCGATGTCCTGGCCGCCAA
      Y  S  F  Q  G  T  Q  T  P  T  E  D  M  R  D  V  L  A  A  K  -

8821 ---------+---------+---------+---------+---------+---------+ 8880
     GCGCATTCCGTATGTCTACATTGATGATGTTGACGGGATTGCTGATGGCCTGGCTGAAGT
      R  I  P  Y  V  Y  I  D  D  V  D  G  I  A  D  G  L  A  E  V  -

8881 ---------+---------+---------+---------+---------+---------+ 8940
     GTTTCGGCTCTACCATCAGGCCCAGGGGCCGGTAGTTATTTTAGCGACGCAAAATGTACT
      F  R  L  Y  H  Q  A  Q  G  P  V  V  I  L  A  T  Q  N  V  L  -

8941 ---------+---------+---------+---------+---------+---------+ 9000
     GGAATCGACTCTCTCGTTGGAGCCGGTGCCGGGTGACTTGCCGCCGGTTTCTGGTTTGCC
      E  S  T  L  S  L  E  P  V  P  G  D  L  P  P  V  S  G  L  P  -

9001 ---------+---------+---------+---------+---------+---------+ 9060
     AGCCTATGACTGCCCACCGATAAGCGACAGTTTTGAGCAAGCCATGGCGCTGATTAATGA
      A  Y  D  C  P  P  I  S  D  S  F  E  Q  A  M  A  L  I  N  E  -

9061 ---------+---------+---------+---------+---------+---------+ 9120
     AGGGCCGGAAAAACTGGTTTGGCAGTTAGGACCGGTCAGTGACGATGAGTATGCCCTGGT
      G  P  E  K  L  V  W  Q  L  G  P  V  S  D  D  E  Y  A  L  V  -
```

(Figure 3.4 continued...)

```
9121 --------+--------+--------+--------+--------+--------+ 9180
     TCATGAAATTGCCGATGCTGCCGGCCTGGCGCTGGTGGATTCGTTAGCGCATCCGGGTTC
      H  E  I  A  D  A  A  G  L  A  L  V  D  S  L  A  H  P  G  S  -

9181 --------+--------+--------+--------+--------+--------+ 9240
     GGCACCGAAATATTATCAGGGCAAGCGCAATCCCCACTATTTGGGAACCCTGGCCATTTA
      A  P  K  Y  Y  Q  G  K  R  N  P  H  Y  L  G  T  L  A  I  Y  -

9241 --------+--------+--------+--------+--------+--------+ 9300
     TGGTTATAGCCCCCGGGTTTACAACTTCCTGCATACCAACGACAAACTCAATCCAATGAG
     G  Y  S  P  R  V  Y  N  F  L  H  T  N  D  K  L  N  P  M  S  -

9301 --------+--------+--------+--------+--------+--------+ 9360
     TGATCAAAGTGTGTTTATGATCAAAAGCCGTGTGGCGCAGATCACCACGCCGTTCTCTGA
      D  Q  S  V  F  M  I  K  S  R  V  A  Q  I  T  T  P  F  S  D  -

9361 --------+--------+--------+--------+--------+--------+ 9420
     TGGTCGGCTCGAGCGCAAGGTGCATTTGGTACAACTCACGCACGATGAACGACACTTATC
      G  R  L  E  R  K  V  H  L  V  Q  L  T  H  D  E  R  H  L  S  -

9421 --------+--------+--------+--------+--------+--------+ 9480
     GCCGTACGCCGATTTGAAACTGCATATGGATTGCCTGACTTTCCTCCGGGCGGTGAAAGC
      P  Y  A  D  L  K  L  H  M  D  C  L  T  F  L  R  A  V  K  A  -

9481 --------+--------+--------+--------+--------+--------+ 9540
     CAATCTGCATGTCGATGCGGCACTACGGGAGAAACGCAAAGCGCTCATCGCTGCCTATCT
      N  L  H  V  D  A  A  L  R  E  K  R  K  A  L  I  A  A  Y  L  -

9541 --------+--------+--------+--------+--------+--------+ 9600
     GGATTCCCCTTCAGATGTTGTTAGCCAGTTACCCAGCCTGCCGATGTCGGCGAACTACTT
      D  S  P  S  D  V  V  S  Q  L  P  S  L  P  M  S  A  N  Y  F  -

9601 --------+--------+--------+--------+--------+--------+ 9660
     TTTTTGCCAATTGAATCGGGTTATCGAGAATTTAATCAAGACGGAAAATTTCGATTTTAC
      F  C  Q  L  N  R  V  I  E  N  L  I  K  T  E  N  F  D  F  T  -

9661 --------+--------+--------+--------+--------+--------+ 9720
     CGGGGTTTACGATGTGGGGCGCTGTGGTATTTCAGCGGTACGCAATGTTGCCAAGACGCG
      G  V  Y  D  V  G  R  C  G  I  S  A  V  R  N  V  A  K  T  R  -

9721 --------+--------+--------+--------+--------+--------+ 9780
     GCGCGGTTTCTCGGGTTGGTATGGTCGGGCCTTGATGGGCGATGCATTACTGGCCACCAG
      R  G  F  S  G  W  Y  G  R  A  L  M  G  D  A  L  L  A  T  S  -

9781 --------+--------+--------+--------+--------+--------+ 9840
     CTATCTGGCGCATACCAGTCCCACCCATGTGGTGGCGTTTATCGGCGACGGGGCCAAAGG
      Y  L  A  H  T  S  P  T  H  V  V  A  F  I  G  D  G  A  K  G  -

9841 --------+--------+--------+--------+--------+--------+ 9900
     GATTGTGCCGGATATTTTACCCGCCTTTATCGACAACATTCTCACCCATCCGCAGTTACT
     I  V  P  D  I  L  P  A  F  I  D  N  I  L  T  H  P  Q  L  L  -

9901 --------+--------+--------+--------+--------+--------+ 9960
     CAATAAAAGCATCACCATTTTCTATTTTTGCAATGGCGGCTTGTCGGTCATCAACACCTA
      N  K  S  I  T  I  F  Y  F  C  N  G  G  L  S  V  I  N  T  Y  -

 9961 --------+--------+--------+--------+--------+--------+ 10020
     TCAGGAACGTATTTTGTTTAACCGCACATCGCGGCAGATGCGCCTGGTGAACGTTGATCA
      Q  E  R  I  L  F  N  R  T  S  R  Q  M  R  L  V  N  V  D  Q  -

10021 --------+--------+--------+--------+--------+--------+ 10080
     ACCCGCGTTTGAACAGACGGTGGATGATTTTCATATTCAGGGTAAAACACTCACTCATTT
      P  A  F  E  Q  T  V  D  D  F  H  I  Q  G  K  T  L  T  H  F  -

10081 --------+--------+--------+--------+--------+--------+ 10140
     TGATGAGGACACCATTCGTCATGCATTGATGACCCCCAAACGACTGAACCTGTTTTCAGT
      D  E  D  T  I  R  H  A  L  M  T  P  K  R  L  N  L  F  S  V  -
```

(Figure 3.4 continued...)

```
10141 ---------+---------+---------+---------+---------+---------+ 10200
      GGTACTGGGGCATAACAATGAAGGGGATGGCATTTCGTTAGCCACAGCCAAAGGCTGGCA
       V  L  G  H  N  N  E  G  D  G  I  S  L  A  T  A  K  G  W  Q  -

10201 ---------+---------+---------+---------+---------+---------+ 10260
      ACGCGACCCCAGCGACCGGGAAGCGCTGCAAGAGCGTAAAGACTGGGCGGCTCGCCAGCC
       R  D  P  S  D  R  E  A  L  Q  E  R  K  D  W  A  A  R  Q  P  -

                                                S-D
10261 ---------+---------+---------+---------AGGAgG-----+---------+ 10320
      CGAATCGACAAGCACTTCATTCGACCAAGGTCAAAACAAGGAAGCTATATCATGAAGTTT
       E  S  T  S  T  S  F  D  Q  G  Q  N  K  E  A  I  S  *
                                                         M  K  F  -
                                                        orf5→

10321 ---------+---------+---------+---------+---------+---------+ 10380
      GGATTTATCGCTCATCCAACCTCGGTTGGTTTAAAACGCTATGTCAAAATGATTGATTTA
       G  F  I  A  H  P  T  S  V  G  L  K  R  Y  V  K  M  I  D  L  -

10381 ---------+---------+---------+---------+---------+---------+ 10440
      TTACAGCGTAATTCGACGGAGTTGCATAGCGGGTACAAACGAGACCTTTGGCGCCGGGAA
       L  Q  R  N  S  T  E  L  H  S  G  Y  K  R  D  L  W  R  R  E  -

10441 ---------+---------+---------+---------+---------+---------+ 10500
      AACCTGGTCCCTTTCATGAACTTTGCCAAAATTACCTCAGCGACCGGCGCTACCTGTGAG
       N  L  V  P  F  M  N  F  A  K  I  T  S  A  T  G  A  T  C  E  -

10501 ---------+---------+---------+---------+---------+---------+ 10560
      GGAGTGATCAAATACATGCCGCTGGTCGCGGATGAGATGCTGGCTGATGCTCGCGGCATT
       G  V  I  K  Y  M  P  L  V  A  D  E  M  L  A  D  A  R  G  I  -

10561 ---------+---------+---------+---------+---------+---------+ 10620
      GCCAATCGTGTGGTGTCGGGTATCGAAGAACTGGTCGAGGATGGCGCTGAACTGGTCGGT
       A  N  R  V  V  S  G  I  E  E  L  V  E  D  G  A  E  L  V  G  -

10621 ---------+---------+---------+---------+---------+---------+ 10680
      TTGGGAGGATTCACCTCCATTGTCGGGCGGCGTGGTGAAGCCACCGCCGAGAAATCACCG
       L  G  G  F  T  S  I  V  G  R  R  G  E  A  T  A  E  K  S  P  -

10681 ---------+---------+---------+---------+---------+---------+ 10740
      GTTCCGGTGACCTCCGGTAATTCGTTAACCACCTATGCGGGTTACAAAGCGCTGATGCAG
       V  P  V  T  S  G  N  S  L  T  T  Y  A  G  Y  K  A  L  M  Q  -

10741 ---------+---------+---------+---------+---------+---------+ 10800
      ATTCAGTCCTGGCTGGATATTCAGCCAGAGCAGGAGCCGGTTGCGATTGTCGGCTATCCG
       I  Q  S  W  L  D  I  Q  P  E  Q  E  P  V  A  I  V  G  Y  P  -

10801 ---------+---------+---------+---------+---------+---------+ 10860
      GGCTCCATCTGTTTGGCGCTGAGCCGGTTATTGCTCGCCCAGGGTTTTTCCCTGCATTTG
       G  S  I  C  L  A  L  S  R  L  L  L  A  Q  G  F  S  L  H  L  -

10861 ---------+---------+---------+---------+---------+---------+ 10920
      CTGCATCGGGCCGGCCATAAAGATGAAGACGAATTGCTCAGCCATTTGCCGGAGCAATAT
       L  H  R  A  G  H  K  D  E  D  E  L  L  S  H  L  P  E  Q  Y  -

10921 ---------+---------+---------+---------+---------+---------+ 10980
      CGCTCACGCGTCACGTTGACCAGTGATCCCGAGGATTTATACCCGCGTTGTAAACTGTTT
       R  S  R  V  T  L  T  S  D  P  E  D  L  Y  P  R  C  K  L  F  -

10981 ---------+---------+---------+---------+---------+---------+ 11040
      GTCGCCGCAACATCCGCCGGAGGCGTGATTGACCCGTATAAACTGCAACCGGGTTCGGTC
       V  A  A  T  S  A  G  G  V  I  D  P  Y  K  L  Q  P  G  S  V  -

11041 ---------+---------+---------+---------+---------+---------+ 11100
      TTTATTGATGTGGCACTGCCCAGAGATATCAATTCTGACACGCGTCCCGATCGGGACGAC
       F  I  D  V  A  L  P  R  D  I  N  S  D  T  R  P  D  R  D  D  -
```

(Figure 3.4 continued...)

```
11101 ---------+---------+---------+---------+---------+---------+ 11160
      ATTTTGATTATCGACGGCGGCTGCGTTACGGCCACCGATGCGGTCAAACTGGGTGGTGAG
      I  L  I  I  D  G  G  C  V  T  A  T  D  A  V  K  L  G  G  E  -

11161 ---------+---------+---------+---------+---------+---------+ 11220
      TCACTGAACGTCACCATTAAACAGCAATTGAATGGCTGTATGGCGGAAACCATTGTTTTG
      S  L  N  V  T  I  K  Q  Q  L  N  G  C  M  A  E  T  I  V  L  -

11221 ---------+---------+---------+---------+---------+---------+ 11280
      GCACTGGAAAATCGCCGGGAGAATTTCTCGTTGGGCCGCTATCTGGCGCTGGATAACGTG
      A  L  E  N  R  R  E  N  F  S  L  G  R  Y  L  A  L  D  N  V  -

11281 ---------+---------+---------+---------+---------+---------+ 11340
      CTTGAAATCGGGGAACTTGCGGAAAAACATGGTTTCCTGGTCTATCCACTGGCCTCTTAT
      L  E  I  G  E  L  A  E  K  H  G  F  L  V  Y  P  L  A  S  Y  -

11341 ---------+---------+---------+---------+---------+---------+ 11400
      GGTGAGCGGATTGATCGTCAACGGGTGATCAATCTCAAGCGTTACTATCACCATGATATT
      G  E  R  I  D  R  Q  R  V  I  N  L  K  R  Y  Y  H  H  D  I  -

11401 ---------+---------+---------+---------+---------+---------+ 11460
      TATTCCGATGAACCTGATACTGAACAACCGCCGGCGTCGCAGTTAGCCTTTATCGACGCG
      Y  S  D  E  P  D  T  E  Q  P  P  A  S  Q  L  A  F  I  D  A  -

                BamHI
                  ↓
11461 ---------+---------+---------+---------+---------+---------+ 11520
      ATTATTGCTCAGGATCCCGCCAGAGAAGACACTCTGGATCGCTATCATCAGTTCATTAAC
      I  I  A  Q  D  P  A  R  E  D  T  L  D  R  Y  H  Q  F  I  N  -

11521 ---------+---------+---------+---------+---------+---------+ 11580
      CCGATGATGGTGGAATTTCTCAAACTGCAACACTGTGACAATGTATTCCGGCGGGCCTCG
      P  M  M  V  E  F  L  K  L  Q  H  C  D  N  V  F  R  R  A  S  -

11581 ---------+---------+---------+---------+---------+---------+ 11640
      GGTACTCAGCTGTTTACCGCTGACGGCGAAGCCTTTCTGGATATGGTGGCCGGTTATGGT
      G  T  Q  L  F  T  A  D  G  E  A  F  L  D  M  V  A  G  Y  G  -

11641 ---------+---------+---------+---------+---------+---------+ 11700
      TGCATCAATCTGGGTCACAATCCGCAACCGATCATCGATGCTTTAAAAGCCTATCTCGAT
      C  I  N  L  G  H  N  P  Q  P  I  I  D  A  L  K  A  Y  L  D  -

11701 ---------+---------+---------+---------+---------+---------+ 11760
      GCTCAGGGGCCCAATTTTATTCAGTATATTTCGATCCCCGAGCAAGCTGCCAAGTTGGCC
      A  Q  G  P  N  F  I  Q  Y  I  S  I  P  E  Q  A  A  K  L  A  -

11761 ---------+---------+---------+---------+---------+---------+ 11820
      GAAGTGCTCTGCCATTTCGCCCCTGGCAATATGGGGAGGGTGTTCTTCAGTAATTCAGGG
      E  V  L  C  H  F  A  P  G  N  M  G  R  V  F  F  S  N  S  G  -

11821 ---------+---------+---------+---------+---------+---------+ 11880
      ACGGAGGCGGTCGAGGCGGCAATGAAGCTGGCCAAAGCGTCGACGGGCAAAGCCGGCATT
      T  E  A  V  E  A  A  M  K  L  A  K  A  S  T  G  K  A  G  I  -

11881 ---------+---------+---------+---------+---------+---------+ 11940
      GCGTATCTGAAAAATAGCTACCATGGCAAAACGCTGGGCGCGTTATCCATTACCGGACGA
      A  Y  L  K  N  S  Y  H  G  K  T  L  G  A  L  S  I  T  G  R  -

11941 ---------+---------+---------+---------+---------+---------+ 12000
      GAAAAACACCGCCGTCATTTTAAGCCACTGCTGGCGTCGATGATTGAAGTACCGTTTGCT
      E  K  H  R  R  H  F  K  P  L  L  A  S  M  I  E  V  P  F  A  -

12001 ---------+---------+---------+---------+---------+---------+ 12060
      GATATTGAGGCACTGCGGCAGACCTTAAGCCGTGATGACATCGGCGCGTTGATGATCGAA
      D  I  E  A  L  R  Q  T  L  S  R  D  D  I  G  A  L  M  I  E  -

12061 ---------+---------+---------+---------+---------+---------+ 12120
      CCGATTCAGGGTGAAGGCGGGGTCCATGTTCCTCCGCCTGGATACCTGCGAACCGTTCAG
      P  I  Q  G  E  G  G  V  H  V  P  P  P  G  Y  L  R  T  V  Q  -
```

(Figure 3.4 continued...)

```
12121 ---------+---------+---------+---------+---------+---------+ 12180
      GAGATCTGCCGTCAAACCGATACCTTACTGATGGTTGACGAAGTGCAAACCGGGTTGGGG
      E  I  C  R  Q  T  D  T  L  L  M  V  D  E  V  Q  T  G  L  G   -

12181 ---------+---------+---------+---------+---------+---------+ 12240
      CGCACCGGGAAGCTCTTTGCCTGTGAGTGGGAAGGGATCGAGCCGGATGTACTGATGCTA
      R  T  G  K  L  F  A  C  E  W  E  G  I  E  P  D  V  L  M  L   -

12241 ---------+---------+---------+---------+---------+---------+ 12300
      TCGAAATCACTGTCCGGCGGGGTAATGCCTATTGGCGCCACGCTATGTCGGGCCATATTT
      S  K  S  L  S  G  G  V  M  P  I  G  A  T  L  C  R  A  I  F   -

12301 ---------+---------+---------+---------+---------+---------+ 12360
      GGCAACGGGCCGTATGGCACCGCAGACCGCTTTTTGATGCACAGCTCGACCTTTGGCGGC
      G  N  G  P  Y  G  T  A  D  R  F  L  M  H  S  S  T  F  G  G   -

12361 ---------+---------+---------+---------+---------+---------+ 12420
      GGGAATATCGCGGCGGTAGTGGCGCTCAGTGCGTTAAGAGAAATTCTGGCTCAGGATCTG
      G  N  I  A  A  V  V  A  L  S  A  L  R  E  I  L  A  Q  D  L   -

12421 ---------+---------+---------+---------+---------+---------+ 12480
      GTCGGGAACGCCGAGCGACTAGGCACGTACTTTAAGCAGGCGCTTACCGATGTCGCTGCC
      V  G  N  A  E  R  L  G  T  Y  F  K  Q  A  L  T  D  V  A  A   -

12481 ---------+---------+---------+---------+---------+---------+ 12540
      CGATACCCCTTTGTGGCGGAAATAGCCGGACGCGGCTTGATGTTAGGGATCCAGTTCGAT
      R  Y  P  F  V  A  E  I  A  G  R  G  L  M  L  G  I  Q  F  D   -

12541 ---------+---------+---------+---------+---------+---------+ 12600
      CAAACCTTCGCTGGTGCCGTAGGTGCTTCGGCCCGGGAGTTTGCCACCCGACTGCCCGGT
      Q  T  F  A  G  A  V  G  A  S  A  R  E  F  A  T  R  L  P  G   -

12601 ---------+---------+---------+---------+---------+---------+ 12660
      GACTGGCACACGACATGGAAATTCCTGCCTGATCCGGTACAAGCCCACTTAAAGGCGGCG
      D  W  H  T  T  W  K  F  L  P  D  P  V  Q  A  H  L  K  A  A   -

12661 ---------+---------+---------+---------+---------+---------+ 12720
      ATGGAGCGTATGGAACAATCACTGGGTGAAATGTTCTGCATGAAATTTGTGACCAAGCTT
      M  E  R  M  E  Q  S  L  G  E  M  F  C  M  K  F  V  T  K  L   -

12721 ---------+---------+---------+---------+---------+---------+ 12780
      TGTCAGGATCACAACATTCTGACCTTTATTACCGCCAACAGCTCAACCGTTATTCGAATT
      C  Q  D  H  N  I  L  T  F  I  T  A  N  S  S  T  V  I  R  I   -

12781 ---------+---------+---------+---------+---------+---------+ 12840
      CAACCGCCACTGACCATCAGCAAGGCTGAGATCGATCGTTTTGTCAGTGCGTTTGCCACG
      Q  P  P  L  T  I  S  K  A  E  I  D  R  F  V  S  A  F  A  T   -

                                          S-D
12841 ---------+---------+---------+--AGGAgG-+---------+---------+ 12900
      GTGTGCGATGAGCTATCAACATTTTTAGAGTAAGGACGCTGAAATGACATTAACCAAGCA
      V  C  D  E  L  S  T  F  L  E  *
                                                M  T  L  T  K  Q   -
                                                orf6→

12901 ---------+---------+---------+---------+---------+---------+ 12960
      AGACGCAGTCAACCAGATGATGGGCTTTTTTCAGTCCAAAACGCTGATCACGGCGCTGTC
       D  A  V  N  Q  M  M  G  F  F  Q  S  K  T  L  I  T  A  L  S
```

one complete and one partial ORF (*orf4* and *orf5*), with the latter apparently translationally coupled to the former. The 5' terminus of *orf5* (coding DNA strand) was deduced when sequence from this study was merged with sequence from the 3' end of the *pig* cluster by A. Cox.

3.4 DNA MOTIFS AND REGULATORY REGIONS

For convenience in this report, the intergenic region of 922nt preceding *orf1* shall be designated intergenic region 1; the region of 163nt between *orf3* and *orf4* shall be designated intergenic region 2. The DNA sequences in intergenic regions 1 and 2 were analysed for areas which might influence transcription or translation.

The nucleotide sequence was examined for translational and transcriptional motifs: Ribosome Binding Sites (RBSs; S-D, after Shine and Dalgarno, 1974) and hexamers resembling the sigma70-type promoter consensus for *E. coli* (Dale, 1994). Additionally, sequence was also analysed for regions resembling alternative (non-σ^{70}) sigma factor-dependent promoters.

3.4.1 RIBOSOME BINDING SITES

Putative RBSs for each of the five ORFs were identified by comparison with the consensus sequence for *E. coli* 16S ribosomal RNA (Dale, 1994b). They are highlighted in Fig.3.4 and also presented separately in Fig. 3.5a. The putative RBS identified for *orf1* compares favourably with the consensus sequence (Fig 3.5a). A good RBS for the first gene of an operon might indicate that the product of this gene is required at the greatest concentration in the cell relative to products of downstream genes in the operon, and often this diminishes the need for RBSs for downstream adjacent genes (Maloy *et al.* 1994a). When sequence upstream of *orf1* was examined for a putative translational start site, it was found that there are three candidate ATG codons present in close proximity to each other within the same frame (Fig. 3.5b). The ATG codon beginning at nt1693 was rejected as a likely start site since upstream sequence does not bear any resemblance to a S-D sequence. The methionine residue encoded from nt1703 was discounted because although a sequence resembling the S-D consensus

FIGURE 3.5 POSSIBLE RIBOSOME BINDING SITES FOR THE PUTATIVE ORFs IN THE 5' END OF THE *pig* CLUSTER

Figure 3.5a

ORF1	AGCAATGGAAGCAA**TGGAGT**GTTTT**ATG**
ORF2	AACTCGGTTTAAG**GGGAAC**ATCATC**ATG**
ORF3	CGGAAAATGTGC**TGGAGG**TGGCGGG**ATG**
ORF4	ATTTTGGCTGCATA**AGGATT**CGGTC**ATG**
ORF5	AAGGTCAAAACA**AGGAAG**CTATATC**ATG**
***E. coli* consensus...AGGAGG.4-7nt.ATG**	

3.5a Designated RBSs and Start Codons.
RBSs are highlighted in red; start codons are highlighted in blue.

Figure 3.5b

```
      ACCGTTGTAGCAATGGAAGCAATGGAGTGTTTTATGGATTTTAACCTGTCAAATAGTCAG
1681  --------+---------+---------+---------+---------+---------+ 1740
                m* e  a  m  e  c  f  M  D  F  N  L  S  N  S  Q  -
                |orf1→
```

3.5b. A Possible Alternative RBS and Start Codon upstream of *orf1*
A candidate hexamer which also resembles a Shine-Dalgarno sequence was found in close proximity to a methionine codon (**m**) upstream of the designated start site (highlighted red) of *orf1*. Also indicated (*) is another methionine residues encoded in the likely leader sequence of *orf1*. See text in section 3.4.1.1 for details.

precedes this codon (GGAAG, Fig.3.5b), it is separated from the accompanying methionine residue by only 2nt, whereas the possible RBS starting at nt1714 is separated from the downstream methionine residue by 4nt. On the basis of spacing, it is therefore, it is proposed that the latter is likely to be the translation initiation point for *orf1*. The putative RBS for *orf2* (gGGAac) contains a GGA motif, but it is nonetheless quite divergent from the consensus. If *orf2* is indeed translationally coupled to *orf1*, there is likely to be a some read-through during translation of the polycistronic message. Indeed, a relatively poor RBS might allow regulation of production of the encoded polypeptide by controlling frequency of translation initiation at this point, i.e. ribosome detachment is more likely to occur in the absence of a good RBS. The proposed RBS for *orf3* (tGGAGG) conforms well to the *E. coli* consensus. It is possible that a good RBS for *orf3* might be required because *orf2* is 2013nt long and would therefore encode a large polypeptide. Since most bacterial mRNAs have a short half life (Dale, 1994b), degradation of such a long message prior to translation of a *sufficient* amount of mRNA for *orf3* could be detrimental to prodigiosin biosynthesis. A good RBS might be present to counter message turnover in this case. The proposed RBS for *orf4* (AGGAtt), conforms to the *E. coli* consensus in only having four out of six residues in common. The putative RBS for *orf5* is at least as good as that of *orf4*, perhaps even better as it is more purine-rich. Once again, this might be to reduce the likelihood of message degradation, since *orf4* is 2600nt in length and would also encode a very large polypeptide.

3.4.3 PROMOTERS

3.4.3.1 PUTATIVE SIGMA-70 TYPE PROMOTERS

The importance of sigma factors in regulation of gene expression was discussed in section 1.4.5.2. The most common promoter sequences found in bacteria are those which interact with sigma 70-RNA polymerase holoenzyme (E σ^{70}) to drive transcription (Dorman, 1993). Transcription of "regular genes" is driven from σ^{70} (RpoD)-dependent promoters, e.g. constitutively expressed genes encoding housekeeping proteins(Ishihama, 1993). However, inducible genes can

also be transcribed from this type of promoter. The nucleotide sequence preceding every putative ORF was examined for regions resembling the *E. coli* σ^{70}-type consensus: TTGACA-17bp-TATAAT (Dale, 1994).

A possible σ^{70} dependent promoter upstream of *orf1* was identified at nt1634 (-35, TTGCAT) and at nt1661 (-10, GATAAT) (Fig. 3.4). The promoter elements are separated by 20nt and the -10 site is approximately 45nt upstream of the translational start site of *orf1*. The sequence and spacing of these elements is therefore suboptimal in comparison to the consensus. A putative promoter upstream of *orf4* was identified at nt7653 (-35, TTGTCT) and at nt7687 (-10, TATATT). Promoter elements are separated by 19nt and the -10 site is located approximately 10nt upstream of the predicted translational start site of *orf4*. Once again, this is suboptimal in comparison to the consensus. No other sequences bearing resemblance to the consensus, in terms of bases and spacing, were found in the 5' end of the *pig* cluster.

3.4.2.2 SCREENING FOR POSSIBLE NON-SIGMA-70 DEPENDENT PROMOTERS

Alternative sigma factors were discussed in section 1.4.5.2. A consensus sequence for RpoN (σ^{54})-dependent promoters has been deduced: CTGGCAC (5bp) TTGCA (6-11bp) +1 (Dorman, 1993). RpoN is structurally and functionally distinct from the RpoD (σ^{70}) family; it is postulated to have more in common with eukaryotic transcription initiation factors than with RpoD (Merrick, 1993). The nonstandard promoter sequence recognised by E-σ^{32} (RpoH), the heat shock sigma factor, is [CTTGAA (13-15bp) CCCCAT-TA (7bp) +1] (Dorman, 1993). The deduced consensus promoter sequence of Eσ^{24} (RpoF)-dependent promoters, responsive to extreme heat shock, is [GAACTT-5bp-ATAAAA-5bp-TCTGA-6bp +1]. As stated in section 1.3.4.2, RpoS (σ^{38}) is considered to be the second most important sigma factor in bacteria. A consensus sequence for RpoS-dependent promoters has not been established. However, studies have shown that some promoters are recognised by both Eσ^{70} and Eσ^{38}, while others are recognised solely by Eσ^{70} or preferentially by Eσ^{38} (Tanaka *et al.*, 1993). Data from base substitution experiments suggest that the -35 region of certain promoters might

function as part of a discriminator mechanism to shift transcription between $E\sigma^{70}$ and $E\sigma^{38}$ (Wise *et al.*, 1996).

No sequences resembling any alternative sigma-factor dependent consensus sequences, in terms of base identity and spacing, were found in the 5' end of the *pig* cluster. This does not exclude the possibility that such factors, in particular RpoS, might be involved in transcriptional regulation of genes in the cluster. It is interesting to note that the putatively assigned -10 region upstream of *orf4* (TATATT) is identical to the putative-10 region of *E. coli osmY*, which is RpoS dependent (Yim *et al.*, 1994; Lange *et al.*, 1993)

3.4.3 TERMINATORS AND REPETITIVE MOTIFS

Intergenic regions 1 and 2 were analysed for other motifs that might be involved in transcriptional or translation regulation, including termination motifs. Sequences were analysed using UWGCG STEMLOOP to identify any areas of DNA resembling a *rho*-independent terminator, i.e., any area of hyphenated dyad symmetry followed by a series of T residues. STEMLOOP identified several palindromic repeats in both intergenic regions and they are shown in Fig. 3.6.

Thirteen possible stem-loop structures were found in intergenic region 1, with a stringency of matching no less than 6 bases. Fig.3.6a shows four inverted repeat sequences 3' proximal region of *orf(-1)*, with two of these (I and II) overlapping with the last two or three codons of the ORF. It is possible that repeat II is a *rho*- independent transcriptional terminator as it contains a G + C-rich region, followed by an A + T-rich region, in which there are two consecutive T residues.

In intergenic region 2 (Fig. 3.6b), 7 indirect repeats were found with a 6-base match stringency. Repeat I is not A + T rich in the in the 3' region as would be predicted for this type of terminator, so this it is unlikely that this region represents a terminator. For any predicted stem-loop structure, the likelihood of

FIGURE 3.6 INDIRECT REPEATS IN THE INTERGENIC REGIONS OF THE '5' END' OF THE *pig* CLUSTER

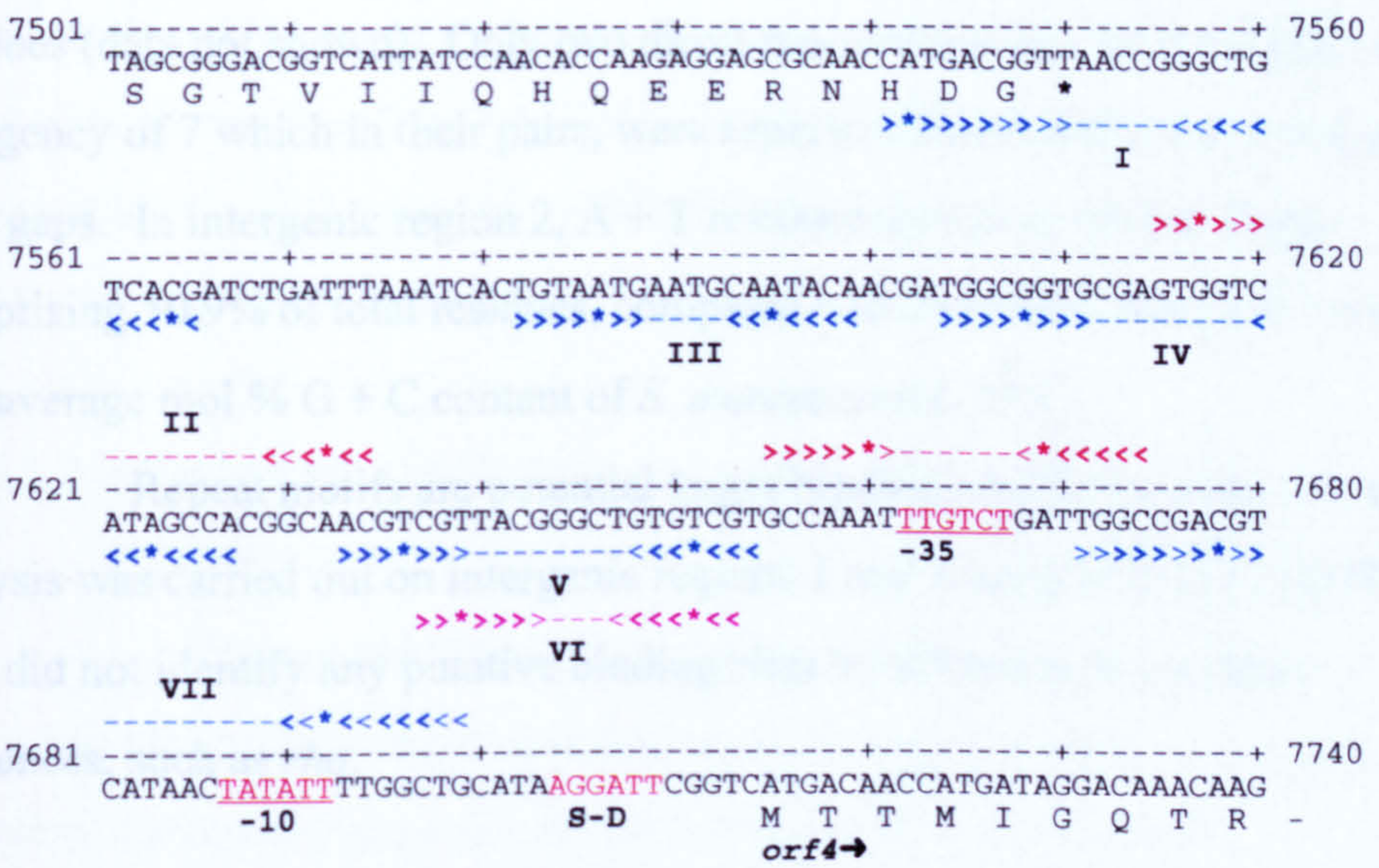
Indirect repeats found using UWGCG STEMLOOP are indicated by chevrons and colour-coded as appropriate for clarity. (*) in a palindrome denotes non-pairing bases; unemboldened chevrons denote nonstandard base pairing. Only the non-coding strand is shown.

Figure 3.6

3.6a Intergenic Region 1, between *orf(-1)* and *orf1*

(Figure 3.6 continued...)

3.6b Intergenic Region 2, between *orf3* and *orf4*



its formation *in-vivo* is dependent upon its free energy (ΔG) (Tinoco *et al.*, 1973). Some of the predicted structures discussed might not therefore be energetically feasible upon transcription in a living system, so prediction of stem-loop structures in a mRNA transcript must be treated with a degree of caution. UWGCG REPEATS and STEMLOOP programs do not assign ΔG values.

Screening both intergenic regions for direct repeats revealed that intergenic region 1 has 17 direct repeats at a base-matching stringency of 7 but no repeats of length greater than 11 matching residues were found (data not shown). It is also noteworthy that intergenic region 1 is abundant in A + T residues. In contrast, intergenic region 2 was found to be devoid of any repeats longer than 8 residues (data not shown). Only two direct repeat sequences were found at a base-stringency of 7 which in their pairs, were separated from each other by 8nt and 92nt gaps. In intergenic region 2, A + T residues are not as predominant, comprising 50.9% of total residues, compared with 66.8% in intergenic region 1. The average mol % G + C content of *S. marcescens* is 58.4.

Repeat motifs are potential target binding sites for proteins, so an analysis was carried out on intergenic regions 1 and 2 using UWGCG MOTIFS. This did not identify any putative binding sites by similarity to database sequences, such as *rho*.

3.5 DISCUSSION AND SUMMARY

3.5.1 OPEN READING FRAMES IN THE 5' END OF THE *pig* CLUSTER

Since prodigiosin is a secondary metabolite, and indeed this study was undertaken to investigate secondary metabolism at the molecular genetic level in this organism, the putative gene arrangement might be of biological significance in this context. The putative gene arrangement may represent one large operon, or two separate transcriptional units . Analysis of sequence data showed that the putative ORFs in the first half of the prodigiosin biosynthetic cluster are separated by a gap of 163 nucleotides. The presence of such an arrangement is suggestive of the possibility that the two units may be subject to differential transcriptional

control, in other words, the two units may not be intrinsically expressed at the same levels, or might be subject to different mechanisms of activation or repression..

Sequencing data suggest that *orf1*, *orf2* and *orf3*, and separately, *orf4* and *orf5,* are translationally coupled by means of overlapping stop/start codons (TN[A]TG or A[TG]A, Fig. 3.4). Translational coupling arrangements are often found in large operons whereby translation of the coupled downstream gene is thought to occur by formation of the 70s initiation complex without prior detachment of the ribosome from the mRNA polycistron. This may aid stabilisation of the message and play a role in ensuring the stoichiometric regulation of the protein products (Gold and Stormo, 1987). Therefore, the close juxtaposition of *orf1, orf2* and *orf3*, and *orf4* and *orf5*, their unidirectional expression, and the possibility of translational coupling suggests that they each may be expressed as an operon.

3.5.2 SUMMARY OF FINDINGS FOR THE 5' END OF THE *pig* CLUSTER

Plasmid pM245 is the smallest functional subclone of pNRT104 which retains the ability to reconstitute pigment production in the *Erwinia carotovora* subsp. *carotovora* heterologous host system. In 11.77 Kb of *S. marcescens pig* cluster DNA which was sequenced in this study, 6 putative ORFs were identified of which two were incomplete: *orf(-1)* is not believed to be part of the *pig* cluster (discussed further in section 4.3.2). A intergenic region of 922bp follows *orf(-1)*, and an intergenic region of 163bp follows *orf3*. Inverted repeats were identified in the intergenic regions. It is possible that one of these, designated repeat II in intergenic region 1, is a *rho*-independent termination motif, free energy permitting.

On the basis of these data alone, it is not possible to ascertain if the putative ORFs in the 5' end of the pig cluster exist as a single operon, or two separate transcriptional units. The presence of a 163bp gap between *orf3* and *orf4* allows the formal possibility that production of encoded proteins might be subject to additional regulation at this point. This possibility seems more attractive in

view of the many factors that are known to impinge upon prodigiosin production in *S. marcescens* (see section 1.3.6), and also because the principal method of controlling gene expression in prokaryotes is by regulating the amount of mRNA produced from genes (Dale, 1994), although this is influenced primarily by the nature of the promoter. The nature of the nucleotide sequence surrounding a promoter can influence its activity, for example by allowing interaction with regulatory proteins which sequester promoter sites from RNA polymerase, or altering the conformation of the strand to reduce promoter efficiency in some way (Dorman, 1994). Seventeen direct repeats were found in intergenic region 1 ,which is also A + T rich. This might be suggestive of possible binding site(s) for regulators, but analysis using computer-based methods failed to find any regulator binding consensus by similarity. However, intergenic region 2 is less A + T rich than intergenic region 1, and contains only two direct repeats of 7nt or more. Intergenic region 2 contains a high number of inverted repeats which could potentially form stem-loop structures, which suggests that this region of DNA may play a regulatory role in expression of downstream prodigiosin biosynthetic genes. As discussed previously in section 1.4.5.2, promoters of secondary metabolite gene clusters are often more complex than those of primary metabolite pathways. The failure to find any immediately obvious motifs which resemble known promoter consensus sequences suggests that the situation in *S. marcescens* is also complex, and although motifs resembling that of *E. coli* σ^{70}-type promoters were found in both intergenic regions of the 5' end of the *pig* cluster, they were not optimal in terms of sequence and spacing for this type of promoter.

3.5.3 THE COMPLETE *pig* CLUSTER SEQUENCE

Merging of data from this study with that obtained by A. Cox (Pers. comm.) revealed that the *pig* cluster is 24.963Kb in length (presented in Appendix IV). A total of 16 putative ORFs were identified in the entire *pig* cluster. They are represented to scale in Fig. 3.7, in which it can be seen that all the putative ORFs, except *orf16,* are unidirectional. It also appears that most of the putatively assigned ORFs overlap one another. The longest intergenic regions are those identified in this study.

Further interpretation of the results presented in this chapter in the context of prodigiosin biosynthesis, requires discussion of data generated from computer-aided analysis of the putative proteins encoded in this region and results of *in-vivo* expression studies. This work is the subject of Chapter 4.

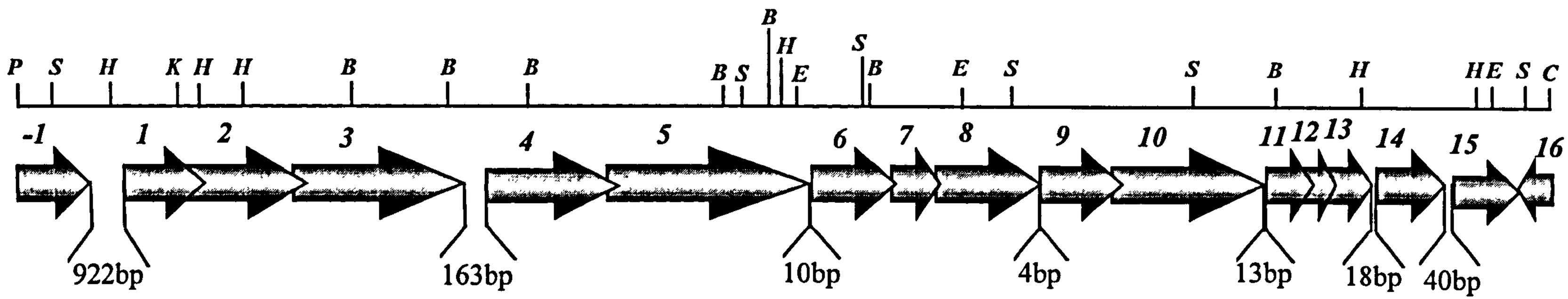


FIGURE 3.7 THE PREDICTED ORFs OF THE *pig* CLUSTER

ORFs are represented by the arrows against the pM245 restriction map and are numberedaccordingly. Intergenic region regions and their sizes are shown where appropriate; all other ORFs have overlapping reading frames. Enzymes Key: ***B***: *Bam*HI; ***C***: *Sac*I; ***E***: *Eco*RI; ***H***: *Hin*dIII; ***K***: *Kpn*I; ***S***: *Sac*I

CHAPTER FOUR

IDENTIFICATION AND ANALYSIS OF THE PUTATIVE PROTEINS ENCODED BY THE 5' END OF THE PRODIGIOSIN BIOSYNTHETIC CLUSTER

4.1.1 PREFACE

Once the complete DNA sequence of the *pig* cluster had been compiled, the putative protein products encoded in the five putative ORFs which had been assigned to the region were analysed further.

4.1.2 AIMS

The overall aim of work described in this chapter was to determine the characteristics of the putative proteins and analyse the potential biological significance of these findings in the context of prodigiosin biosynthesis. This was done in two ways. Firstly, further computer-based analyses were carried out using sequence data from the 5' end of the *pig* cluster, and secondly, the coding regions were expressed *in-vivo* using the T7 system.

4.2.1 FUNDAMENTAL FEATURES OF THE PUTATIVE PROTEINS AND ANALYSES

Translation of the nucleotide sequences of the five ORFs was done using UWGCG MAP and the predicted proteins were then put through a series of computer-aided analyses. For clarity, the features of each putative protein encoded in the 5' end of *pig* cluster DNA shall be presented and discussed in turn, before the final discussion of results.

4.2.1.1 PREDICTED MOLECULAR WEIGHTS AND ISOELECTRIC POINTS

Amino acid translations of the ORFs were analysed using the PREDICT tool at the ExPASy Molecular Biology WWW server at the Swiss Institute for Bioinformatics (SIB, Appendix III). This predicted theoretical molecular weights (MW) and isoelectric point values (pI) for each hypothetical protein.

4.2.1.2 CELLULAR LOCATIONS AND SIGNAL SEQUENCES

Predictions for cellular locations of ORF proteins would be interesting in view of the fact that prodigiosin is a membrane-bound pigment (see section

1.3.1). It is therefore quite plausible that some of the prodigiosin biosynthetic machinery might be located in the periplasm, or might involve integral cell-membrane proteins.

In Gram-negative bacteria, most periplasmic and outer membrane proteins carry a signal sequence at the N-terminus of the protein which is cleaved off after translocation across the cell membrane. The signal is characterised by a net positive charge at the extreme N-terminus adjacent to a hydrophobic "core" region, followed by the signal cleavage region of 5-6 residues (Pugsley, 1993). Additionally, cell-membrane-associated lipoproteins also have characteristic N-terminal signals sequences which are essentially the same as that of export proteins, except that the sequence is usually shorter with a greater proportion of hydrophobic residues in the core region (Pugsley, 1993). Transmembrane regions of cell-membrane associated proteins are usually characterised by a hydrophobic stretch of amino acids at least 20 residues in length in an α-helical conformation (Stryer, 1988a), which is preceded by a strongly positively charged (and therefore, hydrophilic), stretch of residues which is thought to remain in the cytoplasm (von Heijne, 1986), thus anchoring the protein into the membrane (d'Enfert *et al.*, 1989). Computer-based tools which were used to analyse the predicted proteins for hypothetical sorting signals were accessed via the ExPASy Molecular Biology Server at SIB on the WWW: PSORT (prediction of protein sorting signals and cellular locations), and SIGNAL P (prediction of signal peptide cleavage sites). Programs are listed in Appendix III and accompanied by a brief description.

4.2.1.3 PREDICTED SECONDARY STRUCTURES OF PUTATIVE PROTEINS

Prediction of folding of the putative proteins of the 5' end of the *pig* cluster into helices, β-sheets and random coils (i.e., hypothetical secondary structure analysis) was done using UWGCG PROTEINSTRUCTURE and PLOTSTRUCTURE programs (Appendix III). Outputs from these analyses were also compared with separate hydrophobicity plots of the protein sequences which were calculated using the PROTSCALE program (Appendix III) from the ExPASY WWW molecular biology server at SIB.

4.2.1.4 DATABASE SEARCHES AND HOMOLOGUES

The nucleotide sequence of each ORF was entered into BLAST searches (Altschul *et al.*,1990, Gish and States, 1993) on the WWW using BLAST-X (nucleotide sequence translation in 6-frames versus protein sequences) and BLAST-P (protein sequence versus protein sequences) and screened against the non-redundant amino acid database. Homologue information was obtained from SWISSPROT, EMBL, GENBANK and ENTREZ at NCBI, via the WWW as appropriate (Appendix III). For ORFs with significant levels of similarity to database proteins, alignments of amino acid sequences were done using UWGCG PILEUP and PRETTY and percentage values for similarity and identity were ascertained by the use of UWGCG GAP (Appendix III).

4.3.1 THE *Orf(-1)* PREDICTED GENE PRODUCT.

As stated in section 3.3.1, the 5' end of *orf(-1)* is believed to be absent from plasmid pM245. The hypothetical truncate present in pM245 is 262 amino acids long and the predicted MW of the protein translated from the portion of *orf(-1)* present in pM245 is 28853.9 Da, with a predicted pI of 7.8.

4.3.2 PREDICTED CELLULAR LOCATION AND SECONDARY STRUCTURE

As the sequenced region of this protein was thought to be devoid of the N-terminus, any possible signal sequence sites could not be predicted. On the basis of available sequence, PSORT predicts that the protein is located in the cytoplasm. The predicted secondary structure of Orf(-1) is presented in Fig. 4.1

4.3.3 DATABASE HOMOLOGUES OF OF Orf (-1)

BLAST searches revealed that Orf1 is homologous to various bacterial NADH oxidoreductases. It is most similar to NADH oxidase of the sulphur metabolising archaeon *Archaeoglobus fulgidus*. Table 4.1 lists some of the most significantly scoring homologues of Orf(-1) found in the BLAST-P search. A multiple alignment of Orf1 with these homologues is shown in Fig. 4.2

FIGURE 4.1 PREDICTED SECONDARY STRUCTURE OF Orf(-1)

Shown opposite is the "squiggle plot" output from UWGCG PROTEINSTRUCTURE and PLOTSTRUCTURE. Regions of hydrophobicity are represented by diamonds and regions of hydrophilicity by ovals.
Key to symbols:

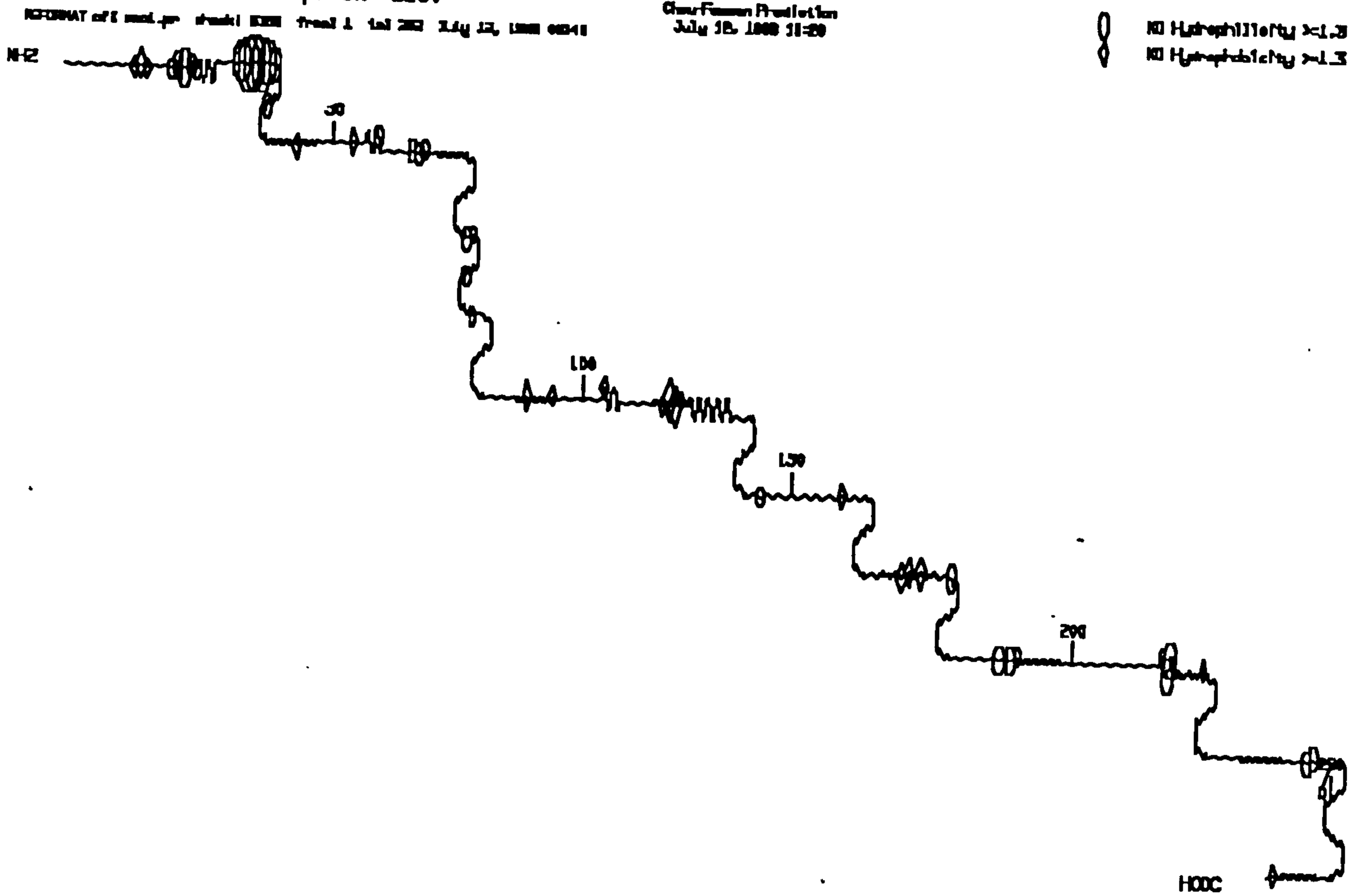
SINE WAVE : α helices

SHARP SAW-TOOTH : β-turns

DULL SAW-TOOTH : random coils

Turns are shown as a 180° turn in the plotting line.

PLOTSTRUCTURE of: emol.pr ck: 8180



Clear Fascan Prediction
July 18, 1988 11:20

NO Hydrophilicity >=1.5
NO Hydrophobicity >=1.5

TABLE 4.1. HOMOLOGUES OF Orf(-1) AS IDENTIFIED IN A BLAST SEARCH AGAINST THE NON-REDUNDANT AMINO ACID DATABASE

HOMOLOGUE	P-SCORE IN BLAST OUTPUT	% IDENTITY NUCLEIC ACID	AMINO ACID (OVERLAP) AND % IDENTITY SIMILARITY	DATABASE, ACCESSION NUMBER AND REFERENCE
NADH oxidase (NOXA-3) *Archaeoglobus fugidus*	$6.8x10^{-47}$	46.3	(260) 36.6 48.2	GENBANK AE001077 Klenk *et al.*, 1997
putative flavoprotein oxidoreductase (PFPOR) *Streptomyces coelicolor* A3(2)	$1.6x10^{-26}$	41.6	(165) 33.3 45.5	EMBL AL023702 Redenbach *et al.*, 1996
Coenzyme A disulphide reductase (COADSR) *Staphylococcus aureus*	$1.7x10^{-21}$	47	(158) 36.5 44.2	GENBANK AF041467 del Cardayre *et al.*, 1998
NADH peroxidase (NAPE) *Enterococcus faecalis*	$7.8x10^{-18}$	37.9	(163) 27.2 40.1	SWISS-PROT P37062 Poole and Clairborne, 1989
NADH oxidase (NAOX) *Enterococcus faecalis*	$1.5x10^{-9}$	34.6	(161) 20.5 29.2	SWISS-PROT P37061 Ross and Claiborne, 1992

FIGURE 4.2 MULTIPLE ALIGNMENT AND CALCULATED CONSENSUS OF Orf(-1) AND ITS HOMOLOGUES

This represents the PRETTY output for Orf1, aligned with the homologues listed in Table 4.1.

Key: *ENTFA* NAPE: *Enterococcus faecalis* NADH peroxidase; *ENTFA* NAOX: *Enterococcus faecalis* NADH oxidase; *ARCFU* NAOX: *Archaeologus fulgidus* NADH oxidase; *STRCO* PFPOR: *Streptomyces coelicolor* A3(2) putative flavoprotein oxidoreductase; *SMA* ORF(-1): *Serratia marcescens* Orf(-1) ; *STAAU* COADSR: *Staphylococcus aureus* coenzymeA disulphide reductase.

Also indicated are proposed NAD and FAD binding sites and the conserved redox-active cysteine. See text in section 4.3.3 for details.

Figure 4.2

```
                  1          |NAD BINDING-|                                50
ENTFA NAPE        ---------- MKViVlGssh gGyeAveell nLhpdAEIqw ..yEkGDFIS
ENTFA NAOX        ---------- MKVVVvGcth AGtSAvksIl anhpeAEvTV ..yErnDnIS
ARCFU NAOX        ---------- MnVVViGGgA AGlkAASrIR RkdgdAsITV V..EaGkyvS
STRCO PFPOR       mnmsagdgks erlVViGGdA AGmSAASQaR RLkgpdElei VafErGhFsS
SMA ORF(-1)       ---------- ---------- ---------- ---------- ----------
STAAU COADSR      ----------m pKiVVvGavA gGatcASQIR RLdkesdI.. iifEkdrdmS
CONSENSUS         ---------- MKVVV-GG-A AG-SAASQIR RL---AEITV V--E-GDFIS

                     Redox-acxtive residue
                  51 ↓                                                   100
ENTFA NAPE        FLSCGmqlYl eGkVKDvnsv RymT.....g EkmeSRGvnV fsnTEiTAIq
ENTFA NAOX        FLSCGialYV GGVVKnaadL fysn.....p EelaSlGatV KMehnVeeIN
ARCFU NAOX        lgrCGlPYYV GGlVhevDnL RetTygavrd Eayfkklkni dvlTEtvAte
STRCO PFPOR       FsaCGiPYwV GGdVtgrDaL iarT.....p EehraRdIdl rMrTEVTeId
SMA ORF(-1)       ---------- ---------- ---------- ---------- ----------
STAAU COADSR      FanCalPYvi GeVVeDrrya layTpekfyd rkq....ItV KtyhEViAIN
Consensus         FLSCG-PYYV GGVVKD-D-L R--T------ E---SRGI-V KM-TEVTAIN

                  101                                                    150
ENTFA NAPE        pkehqVTVKd lvsGeErvEn YDKLiISpGA vPfeldIPGk Dldniylmrg
ENTFA NAOX        VDdktVTaKn lqtGatetvS YDKLVmtTGs wPiiPPIPGI DAenillckN
ARCFU NAOX        iDRsrkTVKi vrnGsEdEln YDyLVIaTGA RPakPPIeGI eAegvvtlts
STRCO PFPOR       VaggrVrard vdsGarswmS YDKLVIgTGA RPvrPelPGI DApgvhgmqt
SMA ORF(-1)       ---------- ---------- ---------- ---------- ----------
STAAU COADSR      deRqtVsVln rktneqfEES YDKLilSpGA ...sanslGf esditftlRN
Consensus         VDR--VTVK- ---G-E-EES YDKLVISTGA RP--PPIPGI DA------RN

                  151                                                    200
ENTFA NAPE        rqwAiklkqK tvDpevnnVV ViGsGYIGiE aAEAfakaGK kVTVIDiLDR
ENTFA NAOX        ysqAnvIIeK aKD..akrVV VVGgGYIGiE lvEAfvesGK qVTlvDgLDR
ARCFU NAOX        aEeAekIIe. mweegaekaV ViGAGfIGlE sAEALknldm EVTVIemmDR
STRCO PFPOR       LdDgqalldt ltrtrgrraV VVGAGYIGvE mAEALinRGy EVTVvn.rgR
SMA ORF(-1        ---------- ---------- ---------- ---------- ----------
STAAU COADSR      LEDtdaIdqf iKanqvdkVl VVGAGYvslE vlEnLneRGl hpTlIhrsDk
Consensus         LEDA--II-K -KD-----VV VVGAGYIG-E -AEAL--RGK EVTVID-LDR

                  201                                                    250
ENTFA NAPE        pLgvYLDKef TDVlteEmEa nnItiatGEt VEryeGD..G rVqkVvTdkn
ENTFA NAOX        ILNKYLDKpf TDVlEkELvd RGvnlalGEn VqqfVaDeqG kVakVITpsq
ARCFU NAOX        vapamLDreM avlvEnhLre kGvnvvtstr VEkIVsqdD. kVRaVIanGK
STRCO PFPOR       epmstLDpdM grmvhgamEg lGItmvndaE Vtgvltgdpg rVRaVaTeda
SMA ORF(-1)       ---------- ---------- ---------- ---------- ----------
STAAU COADSR      I.NKlmDadM nqpildELdk ReIpyrlnEE inaInGne.. ...itfksGK
Consensus         ILNKYLDK-M TDV-E-ELE- RGI----GEE VE-IVGD-DG -VR-VIT-GK

                  251                                                    300
                                                                    |-FAD-BIN-
ENTFA NAPE        aydADlVVVa VGVRPNTawl .KgtLeLhPN GlIktDEYMR T.SePdvfAv
ENTFA NAOX        EfeADmVimc VGfRPNTELl .Kdkvdmlpn GAIeVnEYMq T.SnPdIfAa
ARCFU NAOX        EyPADvVVVa tGikPNsELa EKAGLKiGet GAIwVDEYMR T.SdesIyAG
STRCO PFPOR       EfPADvVVlg iGVRPeTgLa aaAGLpLGah GgltDlaMR vrghenIwAG
SMA ORF(-1)       ---------- ---------- ---------- ---------- ----------
STAAU COADSR      vehyDmiieg VGthPNskfi EssniKLdrk GfIpVndkfe T.nvPnIyAi
Consensus         E-PAD-VVV- VGVRPNTEL- EKAGLKLGPN GAI-VDEYMR T-S-P-I-AG

                     301                                                 350
               -DING|
ENTFA NAPE        GDatlikYnp adTEvnIaLA TNArkQGRfa vkNleepvk. pFPGVQGssg
ENTFA NAOX        GDsavVHYnp sqTknyIPLA TNAvrQGmlV GrNlTeqkl. ayrGtQGTsg
ARCFU NAOX        GDCvEttclV tGkkiiaPfg dvANkQGRvi GENiTGGrA. vFPGVirTAI
STRCO COADSR      GDCvEVldlV sGqErhIPLg ThANkhGqvV GtNvgGGyA. tFPGVvGTAv
SMA ORF(-1)       ------aewV tGnEmlvPLA gpANrQGRia aENmlGGer. ryqrsQGTAI
STAAU COADSR      GDiatsHYrh vdlpasvPLA wgAhraasiV aEqiaGndti eFkGflGnnI
Consensus         GDC-EVHY-V -GTE--IPLA TNAN-QGR-V GEN-TGG-A- -FPGVQGTAI
```

Figure 4.2

```
                    1          |NAD BINDING-|                                50
ENTFA NAPE          ---------- MKViVlGssh gGyeAveell nLhpdAEIqw ..yEkGDFIS
ENTFA NAOX          ---------- MKVVVvGcth AGtSAvksIl anhpeAEvTV ..yErnDnIS
ARCFU NAOX          ---------- MnVVViGGgA AGlkAASrIR RkdgdAsITV V..EaGkyvS
STRCO PFPOR         mnmsagdgks erlVViGGdA AGmSAASQaR RLkgpdElei VafErGhFsS
SMA ORF(-1)         ---------- ---------- ---------- ---------- ----------
STAAU COADSR        ----------m pKiVVvGavA gGatcASQIR RLdkesdI.. iifEkdrdmS
CONSENSUS           ---------- MKVVV-GG-A AG-SAASQIR RL---AEITV V--E-GDFIS

                       Redox-acxtive residue
                    51 ↓                                                   100
ENTFA NAPE          FLSCGmqlYl eGkVKDvnsv RymT.....g EkmeSRGvnV fsnTEiTAIq
ENTFA NAOX          FLSCGialYV GGVVKnaadL fysn.....p EelaSlGatV KMehnVeeIN
ARCFU NAOX          lgrCGlPYYV GGlVhevDnL RetTygavrd Eayfkklkni dvlTEtvAte
STRCO PFPOR         FsaCGiPYwV GGdVtgrDaL iarT.....p EehraRdIdl rMrTEVTeId
SMA ORF(-1)         ---------- ---------- ---------- ---------- ----------
STAAU COADSR        FanCalPYvi GeVVeDrrya layTpekfyd rkq....ItV KtyhEViAIN
Consensus           FLSCG-PYYV GGVVKD-D-L R--T------ E---SRGI-V KM-TEVTAIN

                    101                                                    150
ENTFA NAPE          pkehqVTVKd lvsGeErvEn YDKLiISpGA vPfeldIPGk Dldniylmrg
ENTFA NAOX          VDdktVTaKn lqtGatetvS YDKLVmtTGs wPiiPPIPGI DAenillckN
ARCFU NAOX          iDRsrkTVKi vrnGsEdEln YDyLVIaTGA RPakPPIeGI eAegvvtlts
STRCO PFPOR         VaggrVrard vdsGarswmS YDKLVIgTGA RPvrPelPGI DApgvhgmqt
SMA ORF(-1)         ---------- ---------- ---------- ---------- ----------
STAAU COADSR        deRqtVsVln rktneqfEES YDKLilSpGA ...sanslGf esditftlRN
Consensus           VDR--VTVK- ---G-E-EES YDKLVISTGA RP--PPIPGI DA------RN

                    151                                                    200
ENTFA NAPE          rqwAiklkqK tvDpevnnVV ViGsGYIGiE aAEAfakaGK kVTVIDiLDR
ENTFA NAOX          ysqAnvIIeK aKD..akrVV VVGgGYIGiE lvEAfvesGK qVTlvDgLDR
ARCFU NAOX          aEeAekIIe. mweegaekaV ViGAGfIGlE sAEALknldm EVTVIemmDR
STRCO PFPOR         LdDgqalldt ltrtrgrraV VVGAGYIGvE mAEALinRGy EVTVvn.rgR
SMA ORF(-1          ---------- ---------- ---------- ---------- ----------
STAAU COADSR        LEDtdaIdqf iKanqvdkVl VVGAGYvslE vlEnLneRGl hpTlIhrsDk
Consensus           LEDA--II-K -KD-----VV VVGAGYIG-E -AEAL--RGK EVTVID-LDR

                    201                                                    250
ENTFA NAPE          pLgvYLDKef TDVlteEmEa nnItiatGEt VEryeGD..G rVqkVvTdkn
ENTFA NAOX          ILNKYLDKpf TDVlEkELvd RGvnlalGEn VqqfVaDeqG kVakVITpsq
ARCFU NAOX          vapamLDreM avlvEnhLre kGvnvvtstr VEkIVsqdD. kVRaVIanGK
STRCO PFPOR         epmstLDpdM grmvhgamEg lGItmvndaE Vtgvltgddg rVRaVaTeda
SMA ORF(-1)         ---------- ---------- ---------- ---------- ----------
STAAU COADSR        I.NKlmDadM nqpildELdk ReIpyrlnEE inaInGne.. ...itfksGK
Consensus           ILNKYLDK-M TDV-E-ELE- RGI----GEE VE-IVGD-DG -VR-VIT-GK

                    251                                                    300
                                                                      |-FAD-BIN-
ENTFA NAPE          aydADlVVVa VGVRPNTawl .KgtLeLhPN GlIktDEYMR T.SePdvfAv
ENTFA NAOX          EfeADmVimc VGfRPNTELl .Kdkvdmlpn GAIeVnEYMq T.SnPdIfAa
ARCFU NAOX          EyPADvVVVa tGikPNsELa EKAGLKiGet GAIwVDEYMR T.SdesIyAG
STRCO PFPOR         EfPADvVVlg iGVRPeTgLa aaAGLpLGah GglltDlaMR vrghenIwAG
SMA ORF(-1)         ---------- ---------- ---------- ---------- ----------
STAAU COADSR        vehyDmiieg VGthPNskfi EssniKLdrk GfIpVndkfe T.nvPnIyAi
Consensus           E-PAD-VVV- VGVRPNTEL- EKAGLKLGPN GAI-VDEYMR T-S-P-I-AG

                       301                                                 350
                 -DING|
ENTFA NAPE          GDatlikYnp adTEvnIaLA TNArkQGRfa vkNleepvk. pFPGVQGssg
ENTFA NAOX          GDsavVHYnp sqTknyIPLA TNAvrQGmlV GrNlTeqkl. ayrGtQGTsg
ARCFU NAOX          GDCvEttclV tGkkiiaPfg dvANkQGRvi GENiTGGrA. vFPGVirTAI
STRCO COADSR        GDCvEVldlV sGqErhIPLg ThANkhGqvV GtNvgGGyA. tFPGVvGTAv
SMA ORF(-1)         ------aewV tGnEmlvPLA gpANrQGRia aENmlGGer. ryqrsQGTAI
STAAU COADSR        GDiatsHYrh vdlpasvPLA wgAhraasiV aEqiaGndti eFkGflGnnI
Consensus           GDC-EVHY-V -GTE--IPLA TNAN-QGR-V GEN-TGG-A- -FPGVQGTAI
```

(Figure 4.2 continued...)

```
                351                                                 400
ENTFA NAPE      LaVFDYkfAS TGiNEvMAqk lGketKaVTV vedYlmdfnP dkqkawfKLv
ENTFA NAOX      LylFgwkigS TGVtkesAKl nGLDveatvf EdnYRpefmP ttekvlmeLv
ARCFU NAOX      fKVFDftaAS aGVNEqMAKe aGLDYftVia pSPdRAhYYP qAnyIrlKLi
STRCO PFPOR     sKVcDleiAr TGlrEKdAlR aGLrfvtaTV EStsRAgYYP nASPmtvKml
SMA ORF(-1)     cKVFDhavgS vGlNEKalgR lqmDYqkVyi htPnhAsYYP GsSPIsfKLl
STAAU COADSR    vKfFDYtfAS vGVkpnelKq ..fDYKmVeV tqgahAnYYP GnSPlhlrvy
Consensus       LKVFDY--AS TGVNEKMAKR -GLDYK-VTV ESPYRA-YYP GASPI--KL-

                401                                                 450
ENTFA NAPE      YdpeTtqILG AQlmsKadlt anInaiSlAI QAkMTiEDLA YaDfffqPaF
ENTFA NAOX      YEkGTqRIvG gQlmsKydit qsantLSlAv QnkMTVEDLA isDfffqPhF
ARCFU NAOX      vEkGswRviG AQgVGmgeva KRIDVLStAI QAgMTidqLA nLDLAYAPPy
STRCO PFPOR     aErrTGRlLG vQiVGreGag KRvDiaavAl tAgMTVEqmt aLDLgYAPPF
SMA ORF(-1)     fnpiTGdIfG AQaVGKaGid KcIDViavAq rAhlkVqDLe YLeLAYAPPF
STAAU COADSR    YdtsnrqILr AaaVGKeGad KRIDVLSmAm mnqlTVdeLt efevAYAPPy
Consensus       YE-GTGRILG AQ-VGK-G-- KRIDVLS-AI QA-MTVEDLA YLDLAYAPPF

                451                                                 500
ENTFA NAPE      dkPWniiNTA AleAvkqer~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
ENTFA NAOX      drPWnylNll AqaAlenm~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
ARCFU NAOX      SPalDPViTi AnvAmNkrdG lfeginVfel keklekediv iLDvRseeEf
STRCO PFPOR     SPvWDPVlvA ArKAarever qr~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
SMA ORF(-1)     nsarDvVNqA gmlAsNvinG dtkichVvdi .vnidhqgqc lLDiRtsqEl
STAAU COADSR    ShPkDliNmi gyKAk~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Consensus       SPPWDPVNTA A-KA-N---G ------V--- ---------- -LD-R---E-

                501                                                 550
ENTFA NAPE      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
ENTFA NAOX      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
ARCFU NAOX      KTrriesEkv iHIPilELRe RLdEiPrDKE IvvvCaiGLR sfeAsRiLkh
STRCO PFPOR     ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
SMA ORF(-1)     KTigtypE.a lHIPvdELRg RLnElPkDKE IligCqsGLR ghvAyRlLtq
STAAU COADSR    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Consensus       KT-----E-- -HIP--ELR- RL-E-P-DKE I---C--GLR ---A-R-L--

                551                  571
ENTFA  NAPE     ~~~~~~~~~~ ~~~~~~~~~~
ENTFA NAOX      ~~~~~~~~~~ ~~~~~~~~~~
ARCFU NAOX      aGFekvkile Ggmafwf~~~~
STRCO PFPOR     ~~~~~~~~~~ ~~~~~~~~~~
SMA ORF(-1)     hGFrarnlsg Gyktysltlsa
STAAU COADSR    ~~~~~~~~~~ ~~~~~~~~~~
Consensus     -GF------- G----------
```

Table 4.1 shows that there is a good degree of identity and similarity between Orf(-1) and its homologues. Four of the five most similar proteins are bacterial oxidase or oxidoreductase enzymes. An interesting finding is that Orf1 shares homology with both NADH peroxidase (EC 1.11.1.1) and NADH oxidase (EC 1.6.99.3) enzymes of *Enterococcus faecalis*. Ross and Clairborne (1992) propose that these two flavoproteins constitute a distinct class of FAD-dependent oxidoreductases, namely the flavoprotein peroxide reductases. When the redox centre of *Ent. faecalis* NADH peroxidase was biochemically characterised, it represented the only known example of a flavin-dependent peroxidase (Poole and Clairborne, 1989).

The common feature shared by the enzymes listed in Table 4.1 is that they use, or are predicted to use on the basis of similarity, FAD as a co-factor. Well characterised examples of this family include bacterial D-nicotine oxidase (E.C. 1.5.3.6, Brandsch *et al.*, 1987), and plant reticuline oxidase (E.C. 1.5.3.9, Dittrich and Kutchan, 1991). The NADH oxidase of the *A. fulgidus* was identified by similarity, on the basis of sequence data from the *A. fulgidus* genome sequencing project (Klenk *et al.*, 1997). This is also true of the *Str. coelicolor* A3(2) probable flavoprotein reductase which was identified as such on the basis of similarity to known flavoprotein reductases; the ORF encoding *Str. coelicolor* A3(2) probable flavoprotein oxidoreductase lies in close proximity, to known *red* genes encoding undecylprodigiosin biosynthesis in the chromosome (database accession EMBL AL023702). Coenzyme A disulphide reductase of *Staphylococcus aureus* forms part of the primary thiol/disulphide reductase system in this organism and is proposed to be a member of a novel subfamily of pyridine nucleotide disulphide reductases (delCardayre *et al.*, 1998, in press).

It can be seen in Fig. 4.2 that each homologue contains a conserved cysteine approximately 40 residues in from the N-terminus. In NADH peroxidase of *Ent. faecalis*, this residue is proposed to be the non-flavin redox centre (Poole and Clairborne, 1989). An area which shows good conservation at the extreme N-termini is the probable FAD binding site, by analogy with that which has been identified in *Ent. faecalis* NADH peroxidase (Ross and Clairborne, 1992, Ahmed and Clairborne, 1989). Another fairly well conserved region further down is

proposed, on the basis of similarity, to be the NAD binding site in NADH oxidase of *Ent. faecalis* (Ross and Clairborne, 1992, Ahmed and Clairborne, 1989). Since Orf(-1) aligns with its homologues from approximately residue 300 in 4.3.2, i.e. in the C-terminal region, any potential similarities with the conserved regions discussed above cannot be determined. However, the alignment provides evidence that only a C-terminal portion of an encoded protein is present in the *S. marcescens* DNA in pM245. In order to confirm this, further sequencing on plamid pM245 could be done using oligonucleotides designed from the known DNA sequence in the *orf(-1)* region.

It can also be seen in Fig. 4.2 that the predicted *S. marcescens* protein is ~100 residues longer than the other proteins (with the exception of NADH oxidase of *A. fulgidus*). This suggests that in addition to any catalytic activity it might share with these enzymes, it might have an additional function *in vivo*. The potential biological significance of a longer C-terminal portion in *A. fulgidus* has not, at present, been broached by the workers who identified it. There is also a possibility that the extended C-terminal region of Orf(-1) is actually another ORF. In Fig. 4.2, it can be seen that it has many residues in common with the extended C-terminal region of the *A. fuldigus* protein (without any gaps). This suggests that this is not a sequencing artefact caused by, for example, an undetected frameshift. Despite the fact that that Orf(-1) is not thought to be involved in prodigiosin biosynthesis, it would be interesting to see whether the putative N-terminal portion of the protein is conserved in the same way in *S. marcescens*.

4.4.1 THE *Orf1* PREDICTED GENE PRODUCT

The predicted product of *orf1*, which is hypothesised to be the first gene of the *pig* cluster is 386 amino acids in length and has a predicted MW of 42048.4 Da and a pI value of 5.4.

4.4.2 PREDICTED CELLULAR LOCATION AND SECONDARY STRUCTURE

PSORT analysis of the sequence predicts that Orf1 does not have a cleavable N-terminal signal sequence and no lipoprotein signal sequence.

FIGURE 4.3 PREDICTED SECONDARY STRUCTURE OF Orf1
Shown opposite is the "squiggle plot" output from UWGCG PROTEINSTRUCTURE and PLOTSTRUCTURE. Regions of hydrophobicity are represented by diamonds and regions of hydrophilicity by ovals.
Key to symbols:

SINE WAVE : α helices

SHARP SAW-TOOTH : β-turns

DULL SAW-TOOTH : random coils

Turns are shown as a 180° turn in the plotting line.

PLOTSTRUCTURE of: crf1.pr ck: 2083

However, it is predicted to have a transmembrane spanning domain between residues 82-101. Lack of a cleavable N-terminal signal sequence was confirmed by SIGNALP analysis.

The secondary structure prediction as a "squiggle plot" is shown in Fig. 4.3, where it can be seen that the N-terminus of the structure is not hydrophobic which is therefore in keeping with the prediction that a signal sequence is not present. The hydropathy profile of this region was confirmed by comparison with the PROTSCALE hydrophobicity plot (data not shown). It can be seen in Fig. 4.3, that the region predicted to be a potential transmembrane domain is hydrophobic (confirmed by PROTSCALE, data not shown), but according to these prediction criteria, it is not α-helical in structure and there is no obvious region of hydrophilicity preceding it. Therefore, it is difficult to predict with accuracy, whether a potential transmembrane domain is present on the basis of these findings alone, and if the putative protein would be cytoplasmic or cell-membrane bound.

4.4.3 DATABASE HOMOLOGUES OF Orf1

BLAST searches screening the non-redundant amino acid database revealed that Orf1 has homology to prokaryotic and eukaryotic acyl-CoA dehydrogenases. Orf1 is most similar to RedW of *Str. coelicolor* A3(2), which is predicted by similarity to be an acyl-CoA dehydrogenase (database accession GENBANK 285365). *redW* maps to the *red* locus of *Str. coelicolor* A3(2), encoding the production of undecylprodigiosin. Overall, the highest scoring BLAST hits were bacterial short-chain acyl-CoA dehydrogenases, and examples of these are listed in Table 4.2. Also included for comparison in Table 4.2, is human isovaleryl-CoA dehydrogenase which also showed a significant degree of similarity to the Orf1 sequence. Fig. 4.4 shows alignments of *S. marcescens* Orf1 with its bacterial homologues.

Acyl-CoA dehydrogenases are FAD flavoproteins which catalyse the reduction of acyl-CoA esters in fatty acid oxidation and transfer the electrons to ETF, the electron transfer protein (Stryer, 1988b); they are also involved in the metabolism of branched-chain amino acids (Battaile *et al.*, 1996). Different acyl-

TABLE 4.2 HOMOLOGUES OF Orf1 AS IDENTIFIED IN A BLAST SEARCH AGAINST THE NON-REDUNDANT AMINO ACID DATABASE

HOMOLOGUE	P-SCORE IN BLAST OUTPUT	% IDENTITY NUCLEIC ACID	AMINO ACID (OVERLAP) AND % IDENTITY SIMILARITY	DATABASE, ACCESSION NUMBER AND REFERENCE
RedW *Streptomyces coelicolor* A3(2)	$3.6x10^{-93}$	46.7	(386) 42.1 52.4	GENBANK 2815365 Redenbach *et al.*, 1996; Oliver and Harris, Parkhill *et al.*, direct database submisson
acyl-CoA dehydrogenase A (short chain) (ACDA) *Bacillus subtilis*	$7.0x10^{-61}$	44	(381) 33.8 44.9	SWISS-PROT P45867 Glaser *et al.*, 1995, direct database submission`
short chain acyl-CoA dehydrogenase (SCAD) *Megasphaera elsdenii*	$1.9x10^{-59}$	44	(381) 32.5 43.3	SWISS-PROT Q06319 Becker *et al.*, 1993
isovaleryl-CoA dehydrogenase (IVD) *Homo sapiens*	$4.2x10^{-48}$	44.4	(384) 32.7 43.2	GENBANK 3212539 Tiffany *et al.*, 1997

FIGURE 4.4 MULTIPLE ALIGNMENT AND CALCULATED CONSENSUS OF Orf1 AND ITS HOMOLOGUES

This represents the PRETTY output for Orf1 aligned with the bacterial homologues listed in Table 4.2.

Key: *BACSU* ACDA: *Bacillus subtilis* short-chain acyl-CoA dehydrogenase A; *CLOAB* ACDS: *Clostridium acetobutylicum* short-chain acyl-CoA dehydrogenase; *MEGEL* ACDS: *Megaspaera elsdenii* short-chain acyl-CoA dehydrogenase; *SMA* ORF1: *Serratia marcescens* Orf1; *STRCO* REDW: *Streptomyces coelicolor* probable acyl-CoA dehydrogenase.

Also shown in **bold type** are the two conserved signature regions which identify this family of enzymes. The proposed active site residue critical for catalysis is marked by an arrow and the different residues which may be present in the conserved regions are shown below the relevant positions above or below the alignment. See text in section 4.4.3 for details.

Figure 4.4

```
                1                                                    50
BACSU ACDA      ~~~~MnFsLs eEhemiRklV rdFAkhEvaP tAaERDeqer fdrELfrema
CLOAB ACDS      ~~~~MDFNLt rEQelvRqmV reFAenEvkP iAaEiDeter fpmEnvkkmg
MEGEL ACDS      ~~~~MDFNLt diQqdflkla hdFgekkLaP tvtERDhkgi ydkELidell
SMA ORF1        ~~~~MDFNLs nsQsdiyesa yrFAcdvLdq dAqtRisqki lstELwkkaa
STRCO REDW      mnfdfDagfd tEtrelRdmV vrFArrELds sgrfdDaedf ..rrrwllag
Consensus       ----MDFNL- -EQ---R--V --FA--EL-P -A-ERD---- ---EL-----

                51                                                  100
BACSU ACDA      nlGlTGiPwp edYGGi...G sDyLaYviAV EELSKvcast gvtLSAHiSL
CLOAB ACDS      qyGmmGiPfs keYGGa...G gDvLsYIiAV EELSKvcgtt gviLSAHtSL
MEGEL ACDS      slGiTGayfe ekYGGSgddG gDvLsYIlAV EELaKyDaGv aitLSAtvSL
SMA ORF1        ayGfahgPvs hqfGGSelga lDtalmIeAl gkgSr.DiGl sfsLcAHlca
STRCO REDW      kqGlTGttvp geYGGSglda vsaaatmeAl gygca.DtGf afsvaAHlfa
Consensus       --G-TG-P-- --YGGS---G -D-L-YI-AV EELSK-D-G- ---LSAH-SL

                                                 L    G
                                                GI    S       G
                                                AVS   A S     S
                101           signature sequence: CMTExxNGTD xxA     150
BACSU ACDA      CswPlfaFGT EEQKteYLtq LalGEKIGAf ALTEagsGSD AgsmkTtAer
CLOAB ACDS      CaslInehGT EEQKqKYLvp LakGEKIGAy gLTEPnAGtD sgaqQTvAvl
MEGEL ACDS      CanPIwqFGT EaQKeKfLvp LveGtKlGAf gLTEPnAGtD AsgqQTiAtk
SMA ORF1        CviPlyrFGs sElKdKYLes LvtGklIaAn AaTEPdAGSD iynmQatAqp
STRCO REDW      avmPIveFGT gEQraawLpa LcsGErIaAh AiTEPeAGSD AlhlrTrArp
Consensus       C--PI--FGT EEQK-KYL-- L--GEKIGA- ALTEP-AGSD A---QT-A--

                151                                                 200
BACSU ACDA      igD.dYVLNG SKvFITNGGV ADiyIVFAvT DPeKkkkGvT AFIvEKdfeG
CLOAB ACDS      egD.hYViNG SKIFITNGGV ADtFviFAmT DrtKGtkGis AFIiEKGfkG
MEGEL ACDS      ndDGtYtLNG SKIFITNGGa ADiyIVFAmT DksKGnhGiT AFIlEdGTPG
SMA ORF1        cegG.YiLNG kKIFITNapi ADvFIiyAkT nPdhGflGvs AFliEKGTPG
STRCO REDW      vdDG.hVLsG SKcFITNapV ADvFvVqAaT DPrgGffGlT tFlvEasTPG
Consensus       --DG-YVLNG SKIFITNGGV AD-FIVFA-T DP-KG--G-T AFI-EKGTPG

                201                                                 250
BACSU ACDA      FftGKkEkKl GIRsSPTtEi mFeDcvVPAs krLGeEGeGF kIAMkTLDGG
CLOAB ACDS      FsiGKvEqKl GIRaSsTtEl VFeDmiVPve nmiGkEGkGF pIAMkTLDGG
MEGEL ACDS      FtyGKkEdKm GIhtSqTmEl VFqDvkVPAe nmLGeEGkGF kIAMmTLDGG
SMA ORF1        lnvGevipKd clsncPwsEi VFnDifiPqs qriGmEGaGg aIfhdsmiwe
STRCO REDW      ltvGrpydKv GlRgSPTadv hFdDcyVPAg avLGaEGsGa sIfsssmkwe
Consensus       F--GK-E-K- GIR-SPT-E- VF-D--VPA- --LG-EG-GF -IAM-TLDGG

                251                                                 300
BACSU ACDA      RnGiAAQAvG IAqgALdaAl qYaKeRkQFG KsIaeqQgia fklADMatmI
CLOAB ACDS      RiGiAAQALG IAegAfneAr aYmKeRkQFG rsldkFQgla wmmADMdvaI
MEGEL ACDS      RiGvAAQALG IAeaALadAv eYsKqRvQFG KplckFQsiS fklADMkmqI
SMA ORF1        kgclsAlfvG glarlLettl eYaKaRqQFG KaIgqFQsvS nriiDMklrl
STRCO REDW      RtclfAayLG amrrvLestv dhvrdReQFG spIggFQavS hrivDMlgry
Consensus       R-G-AAQALG IA--AL--A- -Y-K-R-QFG K-I--FQ--S ---ADM---I
```

(Figure 4.4 continued...)

```
                  301                                                  350
BACSU ACDA   EasRLLtYqA AwlessGlPy gkasAmsKLm AgdtAMkVTT eAVQiFGGYG
CLOAB ACDS   EsARyLvYkA AylkqaGlPy tVDaArAKLh AanvAMdVTT kAVQlFGGYG
MEGEL ACDS   EaARnLvYkA AckkqeGkPf tVDaAiAKrv ASdvAMrVTT eAVQiFGGYG
SMA ORF1     EqcRLmlYrA cwkhdqGqda eaDiAmsKLl iSeyAvqsgl dAiQtFGG.a
STRCO REDW   EgARLLlYrA ArslsdGtad eVgpAlAKia vSeaAvqlgl dAVQlrGGlG
Consensus    E-ARLL-Y-A A-----G-P- -VD-A-AKL- AS--AM-VTT -AVQ-FGGYG
                                                signature sequence: QxxGGxG
                                                                    D   S
                                                                    E

                  351                                                  397
BACSU ACDA   .YtkdYPVeR yMRDAKITqI YEGTqEIQrl VIsRmLad-- --------
CLOAB ACDS   .YtkdYPVeR mMRDAKITeI YEGTsEvQKl VIsgkifr-- --------
MEGEL ACDS   .YseEYPVaR hMRDAKITqI YEGTNEvQlm VtggaLlr-- --------
SMA ORF1     amdqElglvR hllnmipsrI fsGTNdIQKe iIaRkLglrg tss-----
STRCO REDW   imdgEaetl. .lRDAlparI fsGTNEIQKn nvaRaLglgr rrpaarr
Consensus    -Y--EYPV-R -MRDAKIT-I YEGTNEIQK- VI-R-L---- -------
(continued...) LxxDxxxxK xxxD                ↑
               I  E    R    E                ACTIVE SITE
               V  N         N                catalytic residue
               M
               F
               Y
```

CoA dehydrogenases have specific catalytic activity for fatty acids of different chain lengths (PROSITE document PDOC0070). In evolutionary terms, it is a highly conserved enzyme family, for example short chain acyl-CoA dehydrogenase of *Met. janaschii* has 44% identity with the human form of the enzyme at the amino acid level (Becker *et al.*, 1993). On the basis of this conservation, and the work which has been done to elucidate the molecular mechanisms of catalysis in this family, two conserved signatures have been found (PROSITE document PDO0070): [GAC]-[LIVM]-[ST]-E-x(2)-[GSAN]-G-[ST]-D-x(2)-[GSA], located in the centre of the enzyme sequence and [QDE]-x(2)-G-[GS]-x-G-[LIVMFY]-x(2)-[DEN]-x(4)-[KR]-x(3)-[DEN], located nearer the C-termini of these proteins.

From the alignments in Fig. 4.4, it can be seen that the degree of conservation of the aligned sequences is very good throughout and the protein lengths are also comparable. The signature sequences, show a very high degree of similarity in the alignments (see Fig. 4.4). However, it can be seen that although this area is completely conserved in database proteins, sequences of *S. marcescens* and *Str.coelicolor* A3(2) deviate from the consensus at certain points. The first of these differences is at residue 127 of the *S. marcescens* sequence, where an alanine residue is present in place of one of isoleucine, leucine, methionine or valine. This change is conserved in the sense that alanine is
also small, neutral and non-polar and since the biological significance of this conserved region is not known, no conclusions can be drawn about this difference on the basis of sequence data alone. The second deviation from the first acyl-CoA signature occurs at residue 137 of the predicted amino acid sequence of Orf1. Here, an asparagine residue is found in place of one of glycine, serine, or alanine. Once again, the biological significance of this change cannot be guessed at on the basis of sequence data alone.

Glutamate active-site catalysis is hypothesised to be initiated by abstraction of the substrate α-proton (Becker *et al.*, 1993). It can be seen in Fig.4.4. that this glutamate residue is conserved in all the aligned proteins at the same relative position *except in the predicted* Orf1 *amino acid sequence*. Here, a conserved change to an aspartate residue is seen. This is potentially a very

important change if this is indeed, by analogy, also the catalytically critical residue in Orf1. In experiments where the active-site glutamate residue of the rat short-chain acyl-CoA dehydrogenase homologue was replaced by asparagine, glycine, glutamine, arginine and lysine, enzyme activity was found to be abolished; replacement with an aspartate residue however resulted in an enzymatically active mutant which displayed reduced catalytic activity (Battaile *et al.*, 1996). In analogous work using recombinant *M. elsdenii* SCAD, replacement of the catalytic glutamate with glycine was shown to result in minimal redox activity (Becker *et al.*, 1993). A further interesting point to note is that the closest Orf1 homologue, RedW, retains glutamate at the same relative position in the alignment.

4.5.1 THE *orf2* PREDICTED GENE PRODUCT

The second putative ORF of the prodigiosin biosynthetic cluster is predicted to encode a protein of 670 amino acids length. The predicted MW is 73860.3Da and it has a theoretical pI of 7.27.

4.5.2 PREDICTED CELLULAR LOCATION AND SECONDARY STRUCTURE

Analysis of the predicted sequence using SIGNALP and PSORT predicts that this protein has a putative N-terminal signal sequence, with the predicted cleavage site after position 19 (AKA↓FP). The predicted secondary structure is shown in Fig. 4.5 It can be seen in Fig. 4.5 that the N- terminal region of the protein is predicted to be hydrophobic, and this was confirmed by comparison with the hydrophobicity plot generated by PROTSCALE (data not shown). PSORT analysis of the Orf2 amino acid sequence predicted that there are three potential transmembrane regions present: the first is predicted to be between residues 53- 71. The second transmembrane region is predicted to between residues 80-96. The third predicted transmembrane region is found between residues 102-118 and the fourth is predicted to be between residues 238-255. These data are supported by hydrophobicity predictions shown on the squiggle plot in Fig. 4.5 and by hydrophobicity plot data generated by PROTSCALE (data

FIGURE 4.5 PREDICTED SECONDARY STRUCTURE OF Orf2
Shown opposite is the "squiggle plot" output from UWGCG PROTEINSTRUCTURE and PLOTSTRUCTURE. Regions of hydrophobicity are represented by diamonds and regions of hydrophilicity by ovals.
Key to symbols

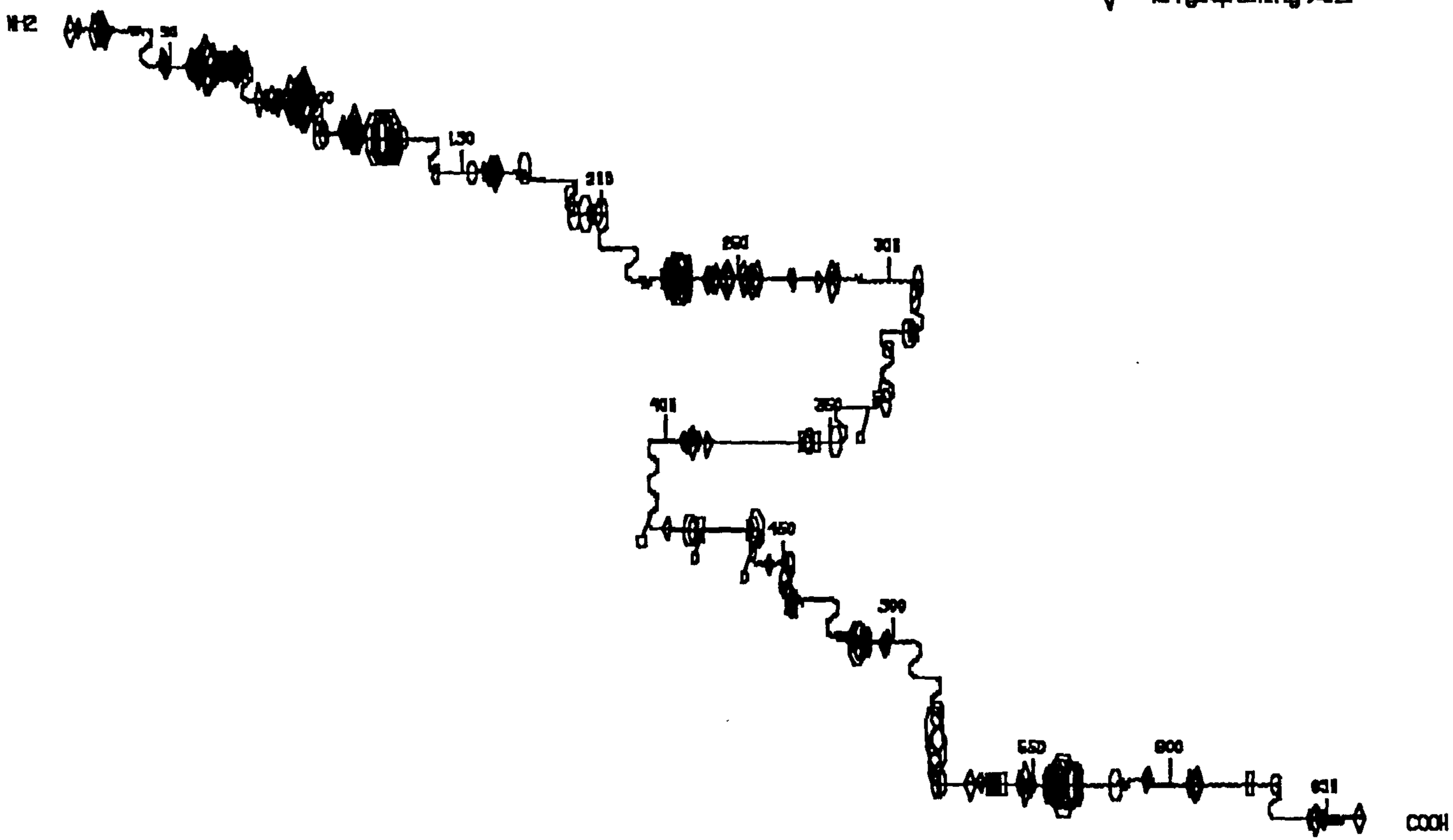
SINE WAVE : α helices

SHARP SAW-TOOTH : β-turns

DULL SAW-TOOTH : random coils

Turns are shown as a 180° turn in the plotting line.

PLOTSTRUCTURE of: orf2.pr ck: 118



not shown), which predicts these regions to be particularly hydrophobic relative to the rest of the protein. However it is difficult to assess, from Fig. 4.5, whether some or all of these regions are also likely to be α-helical, as would be expected for integral cell-membrane domains. Therefore, in consideration of the hydrophobicity data and the presence of a putative signal sequence it can be tentatively predicted that Orf2 is quite likely to be an integral inner-membrane protein with up to four possible transmembrane domains.

4.5.3 DATABASE HOMOLOGUES OF Orf2

BLAST searches of the non-redundant amino acid database for *orf2* and its predicted amino acid sequence showed homology to only one database protein, Hypothetical Protein SC3F7.05c of *Str. coelicolor* A3(2) (EMBL accession AL021409, direct database submission). For convenience, the protein name shall be abbreviated to HYPP in this study.

HYPP is 146 amino acids in length and it maps to cosmid SC357, which carries *Str. coelicolor* A3(2) *red* locus DNA (Redenbach *et al.*, 1996). Orf2 was found to share 43% identity with HYPP at the nucleic acid level, and 40% identity and 50% similarity at the amino acid level in a 144 amino acid overlap. Alignments of the nucleotide and amino acid sequences are shown in Fig. 4.6. As it can be seen in Fig. 4.6, there is homology throughout the overlapping region. Since Orf2 is predicted to be 524 residues longer than HYPP, it is likely that it performs other function(s) in addition to any common function it might have in common with HYPP. It is possible that another protein fulfils any additional functions in the *Str. coelicolor* A3(2) system. HYPP was identified during the *Str. coelicolor* A3(2) genome sequencing project, and in the cognate host, its coding region is located adjacent to *redV* (upstream) and another hypothetical protein (downstream) (database accession EMBL AL021409). Another possible explanation for the differences in length of Orf2 and HYPP is that Orf2 actually represents two proteins which appear as one due to a possible undetected frameshift in the *orf2* DNA sequence, or due to another translational start site in the same reading frame of the corresponding nucleotide sequence. On the evidence of sequence data and analysis this does not seem likely, but

FIGURE 4.6 ALIGNMENT AND CALCULATED CONSENSUS OF THE AMINO ACID SEQUENCES OF Orf2 AND ITS HOMOLOGUE FROM *Streptomyces coelicolor* A3(2)

The consensus sequence is shown below the alignment.

Key: *SMA* ORF2: *Serratia marcescens* Orf2; *STRCO* HYPP: *Streptomyces coelicolor* A3(2) hypothetical protein.

Figure 4.6

```
                1                                                    50
  SMA ORF2      ~~~~~mIiQR lfGiLYmLAG laKaFPQfEn VpavLrqAai ANqGTwyaaa
STRCO HYPP      mtgrnlIlQR vvGaLYiLAG igKfFPQlEs VegrLddAse ANdGTvisgp
  Consensus     ------I-QR --G-LY-LAG --K-FPQ-E- V---L--A-- AN-GT-----

               51                                                   100
  SMA ORF2      siWLgaHgdv inilVgVvlf gsGviLmlNp lwttl.viYa QLLMmavFVV
STRCO HYPP      vdWLdrHptg vmwfVaVamv aaGlaLlwNr rglviaalYg QLLMlvlFVV
  Consensus     --WL--H--- ----V-V--- --G--L--N- --------Y- QLLM---FVV

              101                                                   150
  SMA ORF2      ILhqSqPqvm llDgvFalAA lYmLrgqYHr kPkpRtfpTt sfSlPTPsse
STRCO HYPP      ILvsSvPeil vmDaaFfaAA iYlLy.rYHa sPgaR...Ta apSgPTPegr
  Consensus     IL--S-P--- --D--F--AA -Y-L---YH- -P--R---T- --S-PTP---

              151                                                   200
  SMA ORF2      ssfsaplgde ydvviiggga sgltaaseft hervlvleks stfggnaryh
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------

              201                                                   250
  SMA ORF2      tfnrlkhpta gvcfqepfpg snmlrllkki glegkyksne kdtlvffdtf
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------

              251                                                   300
  SMA ORF2      lllkclgeiv vgfikqpryl lklsvwglts qlflhaiigk pyvvaakqlg
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------

              301                                                   350
  SMA ORF2      dpifadlytf ldkfsprgdf yprlpwtpng swskahmell dnislytylf
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------

              351                                                   400
  SMA ORF2      epdklgrlpe qlrpparlgk lvenavsttl rvecldihdv sayvglhflv
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------

              401                                                   450
  SMA ORF2      gylrgnlvtl pggngsisag lckylshqrn vtlqnhvqlt avepqhngtc
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------

              451                                                   500
  SMA ORF2      iqftingqpr qvqaqqiiwa apktqlatwl pglpakqlaa iknirhedyy
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------
              501                                                   550
  SMA ORF2      lanvflskpv lghsfggymi epdsnkdpfs wckagtclva nwmddhadvd
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------

              551                                                   600
  SMA ORF2      vgvltllkpt trserqdrta qnaflalqqq tyaeiakvlr nigigaevie
STRCO HYPP      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
  Consensus     ---------- ---------- ---------- ---------- ----------
```

(Figure 4.6 continued...)

```
                601                                                   650
  SMA ORF2      diqiwywpag lvtsvvgqqa egvfetarqs fenihfanqd svgvgniesa
STRCO HYPP      ---------- ---------- ---------- ---------- ----------
 Consensus      ---------- ---------- ---------- ---------- ----------

                651                        676
  SMA ORF2      ilsgidaana vkaqlmdten vvevag
STRCO HYPP      ---------- ---------- ------
 Consensus      ---------- ---------- ------
```

unfortunately without further work (discussed in section 4.10), this possibility cannot be discounted either.

As both Orf2 and HYPP map to the *pig* cluster and *red* cluster in their respective hosts, this provides further evidence that these hypothetical proteins are likely to be involved in pigment biosynthesis. Since these proteins do not, at present, have any other database homologues, it can be argued that the production of such distinct metabolites might require hitherto unidentified (and novel) proteins. To test if Orf2 performs some essential role in pigment biosynthesis in *S .marcescens*, this could be verified at the molecular level by site-directed mutagenesis of the encoding region, for example by the construction of a marker-exchange mutant, or by construction of in-frame deletion mutants and assessing the impact this has on the pigment phenotype.

4.6.1 THE *orf3* PREDICTEDGENE PRODUCT

Orf3 is predicted to be 890 amino acids in length and have a molecular weight of 99766.4Da and a theoretical pI of 5.59.

4.6.2 PREDICTED CELLULAR LOCATION AND SECONDARY STRUCTURE

The predicted secondary structure of Orf3 is shown in Fig.4.7. SignalP and PSORT did not predict a N-terminal signal sequence, but the protein PSORT analysis predicts that Orf3 possesses a single transmembrane region between residues 159-177. This is in agreement with PROTSCALE hydrophobicity plot data (not shown) and Fig. 4.7 shows that this region is hydrophobic in this region.

4.6.3 DATABASE HOMOLOGUES OF Orf3

BLAST searches of the non-redundant amino acid database revealed that Orf3, has homology to various prokaryotic phosphoenol pyruvate (PEP) utilising enzymes. Some of the highest scoring BLAST hits are presented in Table 4.3. BLAST search results showed that Orf3 is most similar to protein SC3F7.16, from *Str. coelicolor* A3(2) which, on the basis of similarity, is hypothesised to be a PEP-utilising enzyme (database acsession GENBANK 2808757). For

FIGURE 4.7 PREDICTED SECONDARY STRUCTURE OF Orf3

Shown opposite is the "squiggle plot" output from UWGCG PROTEINSTRUCTURE and PLOTSTRUCTURE. Regions of hydrophobicity are represented by diamonds and regions of hydrophilicity by ovals.

Key to symbols:

SINE WAVE : α helices

SHARP SAW-TOOTH : β-turns

DULL SAW-TOOTH : random coils

Turns are shown as a 180° turn in the plotting line.

PLOTSTRUCTURE of: orf3.pr ck: 4021

TABLE 4.3 HOMOLOGUES OF Orf3 AS IDENTIFIED IN A BLAST SEARCH AGAINST THE NON-REDUNDANT AMINO ACID DATABASE

HOMOLOGUE	P-SCORE IN BLAST OUTPUT	% IDENTITY NUCLEIC ACID	AMINO ACID (OVERLAP) AND % IDENTITY SIMILARITY	DATABASE, ACCESSION NUMBER AND REFERENCE(S)
SC3F7.16 PEP utilising enzyme (PEPUE) *Streptomyces coelicolor* A3(2)	$4.6x10^{-227}$	49.44	(846) 40.4 50.2	GENBANK 2808757 Redenbach *et al.*, 1996, Oliver and Harris, Parkhill *et al.*, direct database submission
probable PEP synthase (PPSA) *Methanococcus janaschii*	$2.7x10^{-67}$	41.6	(890) 23.2 34.6	SWISS-PROT Q57962 Bult *et al.*, 1996
probable PEP synthase (PPSA) *Staphylothermus marinus*	$1.6x10^{-64}$	40.3	(746) 29.2 40.6	SWISS-PROT P46893 Cicicopol *et al.*, 1994
PEP synthase (PPSA) *Escherichia coli*	$6.1x10^{-42}$	42	(674) 28.3 37.8	SWISS-PROT P23538 Niersbach *et al.*, 1992

convenience, SC3F7.16 shall be designated PEPUE in this study. The ORF encoding PEPUE was identified during the *Str. coelicolor* A3(2) genome sequencing project and it maps to the end of the chromosomal region corresponding to DNA in cosmid SC3F7 (Redenbach *et al.*, 1996), downstream of several known *red* genes and adjacent to a probable methyltransferase-encoding ORF (SC3F7.15, database accession data) and a probable oxidase (SC10A5.02, database accession data). Other database proteins that are homologous to Orf3 include two PEP synthases from thermophilic bacteria (see Table 4.3); a lesser degree of similarity, due to the differences in protein length, was also found to *E. coli* PEP synthase. Alignments of Orf3 and the homologues in Table 4.3 are shown in Fig. 4.8.

PEP utilising enzymes share a common catalytic mechanism: they bind PEP and transfer the phosphoryl group from this metabolic intermediate to a histidine residue, and they have been found to be structurally related in this respect (Reizer *et al.*, 1993, Pocalyko *et al.*, 1990). Examples include the database homologues listed in Table 4.3, which, with the exception of *Str. coelicolor* A3(2) PEPUE, have been experimentally characterised as PEP synthases. Other enzymes in this family include pyruvate orthophosphate dikinase (EC 2.7.9.1) which catalyses the ATP-dependent reversible phosphorylation of pyruvate and phosphate to PEP, and PEP-protein phosphotransferase (EC 2.7.3.9) which is the first enzyme of the PEP-dependent sugar phosphotransferase system (PTS) of bacteria (PROSITE Document PDOC00527). PEP synthases (EC 2.7.9.2) catalyse the ATP-dependent, reversible phosphorylation of pyruvate to PEP, adenosine monophosphate (AMP), and phosphate (PROSITE Document PDOC00527).

From the known PEP utilising enzymes, characteristic signature sequences have been deduced which are found exclusively in this enzyme family, the first of which is located around the catalytic histidine residue: G-[GA]-x-[TN]-x-H-[STA]-[STAV]-[LIVM](2)-[STAV]-[RG]. The second signature sequence whose biological significance is unknown, is located in the C-terminal regions of these enzymes: [DEQS]-x-[LIVMF]-S-[LIVMF]-G-[ST]-N-D-[LIVM]-x-Q-

FIGURE 4.8 MULTIPLE ALIGNMENT AND CALCULATED CONSENSUS OF Orf3 AND ITS HOMOLOGUES

Fig. 4.8a
This represents the PRETTY output for Orf3 aligned with the bacterial homologues listed in Table 4.6. Active site histidine residues are indicated by an arrow.
Key: *SMA* ORF3: *Serratia marcescens* Orf3; *STRCO* PEPUE: *Streptomyces coelicolor* A3(2) PEP utilising enzyme; *ECOLI* PPSA: *Escherichia coli* PEP synthase; *STAMA* PPSA: *Staphylothermus marinus* PEP synthase; *METJA* PPSA: *Methanococcus janaschii* PEP synthase.
Also shown in **bold type** are the two conserved signature regions which identify this family of enzymes. The proposed active site residue critical for catalysis is marked by an arrow and the different residues which may be present in the conserved regions are shown below the relevant positions above or below the. See text in section 4.6.3 for details.

Figure 4.8a

```
                1                                                    50
SMA ORF3        ---------- ---------- ------mnqp lvVei.sgdk alehh.hlGG
STRCO PEPUE     mdthvsdttt pagagegpag pvtdpaGhaa cvVtlgaggp tgghrgelGG
ECOLI PPSA      ---------- ---------- -msN.nGss. plV.lWynql gmnDVdrvGG
STAMA PPSA      ---------- ---------- -msNvpkek. rfi.vWldev tkdDVvlvGG
METJA PPSA      ---------- ---------ml iiqNtkGdsm kfi.aWldel snkDVdiaGG
Consensus       ---------- ---------- ---N--G--- --V--W---- ---DV---GG

               51                                                   100
SMA ORF3        KGySLnnlI. ...hAGLPVP sAFcVTAqAY qqFIeEvvp. ........ga
STRCO PEPUE     KGtrLaEls. ...aAGLPVP PAFclTtalf daylrETGi. ........aa
ECOLI PPSA      KnASLGEMIt nlsgmGvsVP ngFatTAdAf nqFldqsGvn qrIyelLdkt
STAMA PPSA      KnAnLGEMIr ....AGiPVP PgFaVTAyAY kyFIekTGlk dkIyplLnsi
METJA PPSA      KGASLGEM.. ..wnAGLPVP PAFvVTAdAY rhFIkETGlm dkIreiLsgl
Consensus       KGASLGEMI- ----AGLPVP PAF-VTA-AY --FI-ETG-- --I---L---

               101                                                  150
SMA ORF3        eltDgDlia. ....vRdaIl hadiPdsLkq AIgdAYqhLg h.......dt
STRCO PEPUE     eaagaDprt. ....lRerIl gTrMPasiad AvldAYgsmg r........p
ECOLI PPSA      DidDvtqLak agaqIRqwIi dTpfqpeLen AIreAYaqLs ....addEna
STAMA PPSA      DvnDkkvLde ttakIRqwIm dTpMPpevee eIrkyYreLa kkigmepEkl
METJA PPSA      DvnDtDaLtn askkIRklIe eaeMPedLrl AIieAYnkLc ..mcgedE.v
Consensus       D--D-D-L-- ----IR--I- -T-MP--L-- AI--AY--L- -------E--

               151                                                  200
SMA ORF3        tiAVRSSAld EDgqrqSFAG QyeTYLhVkG sEaVLhkVqa CWASLwaeRA
STRCO PEPUE     rVAVRSSglr EDsaaqSFAG QhDTvLdVcG dEdVLdaVlr CWASLwsdRA
ECOLI PPSA      sfAVRSSATA EDmPdASFAG QQeTfLNVqG fdaVLvaVkh vfASLFndRA
STAMA PPSA      rVAVRSSATA EDmPeASFAG QQDTYLNVyG ednVvyyVkr CWASLFtsRA
METJA PPSA      tVAVRSSATA EDlPeASFAG QQDTYLNikG aEnVvkyVqk CfsSLFtpRA
Consensus       -VAVRSSATA ED-P-ASFAG QQDTYLNV-G -E-VL--V-- CWASLF--RA

               201                                                  250
SMA ORF3        aqYR.hesas H..sAiAVil QvMVdaDa.. AGVMFTqDPl SGstdkVVId
STRCO PEPUE     tvYR.dtdap ...dALAVvV QeMihtDv.. sGVMFTvDPv nprphrlVvE
ECOLI PPSA      isYRvhQGyd HrgvALsagV QrMVrsDlas sGVMFsiDte SGfdqvVfIt
STAMA PPSA      vfYRvaQGip HekslmsVtV QkMVns..rt AGVMFTlhPv tGdekvVVIE
METJA PPSA      ifYReqQGfd HfkvALAavV QklVnae..k AGVMFTvnPi SenydelVIE
Consensus       --YR--QG-- H---ALAV-V Q-MV--D--- AGVMFT-DP- SG----VVIE

               251                                                  300
SMA ORF3        scWGLGEGVV SGqVttDsFt lDKaTgElcD qqIrhKpn.. ..........
STRCO PEPUE     AcqGLGEGlV SGqVssDfFv VDdekLEvVe ervryKvt.. ..........
ECOLI PPSA      saWGLGEmVV qGaVnPDeFy VhKpTLaanr paIvrrtmg. ...skk....
STAMA PPSA      AsWGLGEsVV gGkVtPDewv VDKqTLqiVD qkIhhKtlai vfdpkkgknv
METJA PPSA      AaWGLGEGVV SGsVsPDtyi VnKkTLEiVD khIarKetmf vkd.......
Consensus       A-WGLGEGVV SG-V-PD-F- VDK-TLE-VD --I--K---- ----------

               301                                                  350
SMA ORF3        ........yc qrdEhGl.Vt llqtPeakRD lPsLTpaqlq qLvtLArqaq
STRCO PEPUE     ........kc aplEpGr.ig mtkVdaaaRs vPcLThdqlr ELgaLAvrIr
ECOLI PPSA      .irm...vya ptqEhGkqVk iedVPqeqRD ifsLTneEvq ELAkqAvqIE
STAMA PPSA      eirwdenkqa wvsEeG.pVd iemVkhfhpD kPaLkeeEvk rLAeLAllIE
METJA PPSA      .......... ...EkG.etk vveVPddmke kqvLsddEik ELAkiglnIE
Consensus       ---------- ---E-G--V- ---VP---RD -P-LT--E-- ELA-LA--IE

               351                                                  400
SMA ORF3        liYstelDIE WAVk.....d dKvwllQARP vTT..sakta nviyanpw..
STRCO PEPUE     dlYGseqDIE WgVr.....d GvfhlfQtRP iTTrpaapaa psgalspyva
ECOLI PPSA      KHYGrpMDIE WAkd...ght GKlfivQARP eTvrs..... .....rgqvm
STAMA PPSA      KHYGrhMDIE WAVdydipfp dnvfivQARl eTvwsvrkek ekaekkaeik
METJA PPSA      KHYGkpMDvE WAye.....k GKfymlQARP iTT....lkk gkkekka..k
Consensus       KHYG--MDIE WAV------- GK----QARP -TT------- ----------
```

(Figure 4.8a continued...)

```
                401                                                 450
SMA ORF3        .esdpaakEG affsR.mdtG e.IvtGlmtp lglsfcqfyQ khihGpaikT
STRCO PEPUE     pqpepvl.nG tlwsR.mdiG e.IfvGlmtp lglsfaryyQ rnvhtDcaga
ECOLI PPSA      erytlh.sqG KiiaeGraiG hrIgaGpvKv IhD..isEmn riepGDvLVT
STAMA PPSA      gknivklsEa KvlvRGlpas pgIgaGvaKv IfDphskEaQ efkeGevLVT
METJA PPSA      eedi....Ea KillkGigas pgIatGvvKi IhD..vsEid kvkeGDiLVT
Consensus       --------EG K---RG---G --I--G--K- I-D----E-Q ----GD-LVT

                                          active site catalytic residue
                                                  ↓
                451                               500
SMA ORF3        .MglaDishw qiymgyI... qGyvylnisg sAymL..rqc ppTrnemkft
STRCO PEPUE     .lgvrDtgea dlhmgfy... qGhvylnisy ssylL..aqc lpTrdqrhft
ECOLI PPSA      dMTdPDwePi MKKAsAIVTn rGGrTchaai iAreLgipav vgcgdaterm
STAMA PPSA      kMTdPDwvPl MKKAvAIVTd eGGmTshaai vsreLgipai vgTgnatqvi
METJA PPSA      eMTtPDmvPa MKKAaAIVTd eGGlTcieg. dAkiLtdrgf lkmkevyklv
Consensus       -MT-PD--P- MKKA-AIVT- -GG-T----- -A--L----- --T-------
                        signature #1: GGXTXHSSL LSR
                                      A N  TTI ITG
                                           AAV VA
                                            VM MV

                501                                               550
SMA ORF3        tryatdeidf kDyknpygag vqgwdyaksc wywlkqqVrn mrsaartveq
STRCO PEPUE     srfvseeVdl aDyenpfgtf pggmEdllst vhwlqhtare mtqmksrsqq
ECOLI PPSA      KdGenvtVsc aegdt.gyvy aellEfsvks ........ss vetmpd..lp
STAMA PPSA      KsGievtV.. .Dgsr.gvvy egivEdlvkp keevkaeVag vgispeqllp
METJA PPSA      KnGeklkVlg lnaetlktew keiidaqkre arryeigVyr knkntkdtik
Consensus       K-G----V-- -D-------- ----E----- -------V-- ----------

                551                                               600
SMA ORF3        mialrqdett rflGldltam tlqqLdqE.. .LqRIdrFfl dscaaYmPff
STRCO PEPUE     mvdarlyefd rarGldltrm srreLhgE.. .LhRdlawfh dmhigYmPyy
ECOLI PPSA      l.....kvm mNvGnpdraf dfacLpnEgv gLaRle.FIi nrmigvhPra
STAMA PPSA      lypvtatkiy mNlGepdaie kykdLpfdgi gLmRIe.FIi tdwvqYhPly
METJA PPSA      itpdhkfpvf vN.Gelskv. qlcdiidnnl svlsId.yI. .......P..
Consensus       ---------- -N-G------ ----L--E-- -L-RI--FI- -----Y-P--

                601                                               650
SMA ORF3         LqsFalydal aqaceRhi.k dgkGlqnriK aSmnnLrtie vtlGIiklva
STRCO PEPUE      inaFafygll telcaRwlgs dgvGlqnrvK tdmssLrtve sakevwsvaq
ECOLI PPSA       LlEFddqepq lqneiRemmk ...G.....f dSprefyvgr lteGIatlga
STAMA PPSA       LiE..... .........q ...G.....K eS...Lfidk laeGIakvaq
METJA PPSA       miE.....ek yeslaevmyl ...G.....g avlsdghivr .rnG......
Consensus        L-EF------ -----R---- ---G-----K -S---L---- ---GI-----

                651                                               700
SMA ORF3        tvnqqtelka lFeqhradEl vtLlpvhdis rafwqgdfed fLvefGsRGr
STRCO PEPUE     AakndPavlR iikDepledi arLlredpaG qrfwdrhmeP fLranGtRGh
ECOLI PPSA      AfypkrviVR .lsDfksnEy anLv....gG EryepdeenP mL...GfRGa
STAMA PPSA      AiyprPvvVR .FsDfktnEy rgLk....gG EkyepeernP mi...GwRGv
METJA PPSA      ....kPirVR .FtqkdteEk kdfi...... Ekvkgdvk.. li...G....
Consensus       A----P--VR -F-D----E- --L------G E--------P -L---G-RG-

                701                                               750
SMA ORF3        qefdlsiPrw RDdpsyllQv mKmylqhpvd Lhkk.lrete llRqqDs.ea
STRCO PEPUE     qemeithPrw iDdpsyifQm irryvadgfs iddi.lrrss gw.sdDs.re
ECOLI PPSA      gry..vsdsf RDcfalEcea vKrvrnd.mG LtnVei.miP fvRTvDqaka
STAMA PPSA      sry..ihPky epafrlEvra iKkvree.mG LtnVwv.mfP fvRTtweler
METJA PPSA      gnf..ieisn Rnnv.iEyQt srkipseilG fieVnintiP lyaTkDeiad
Consensus       -------P-- RD----E-Q- -K-------G L--V-----P --RT-D----
```

(Figure 4.8a continued...)

```
                751                                                       800
SMA ORF3        lfsamswsGr fklktliKly gMmaerrEat rptfitetwf yrcimleVlr
STRCO PEPUE     vlgrlpmpkr qvldtvisly alcselrEtt rmsmitsiwl vrnvvyeVGr
ECOLI PPSA      vveelarqGl krGenglKii mMc....Eip snallaeqfl eyfdGfsiG.
STAMA PPSA      alkimeeeGl krGkd.fKvw aMa....Evp sivlladkfa eyvdGfsiG.
METJA PPSA      li.....aGf vdGdgclsgk rrv....Eiy qn....sshi kkieGliVG.
Consensus       --------G- --G----K-- -M-----E-- ---------- ----G--VG-
                                                signature #2: DXLSLG
                                                              E I I
                                                                V V
                                                                M M
                                                                F F

                801                                                       850
SMA ORF3         rldaqgiass ad.Lpyvdfe qfrayvagti paeqaFskaR lDqnrhqhlf
STRCO PEPUE      rlvadgvlhs pdeiahLdfe dvrrylagde davrvFDraR iDAarrlheh
ECOLI PPSA       .......snd mtqL.aLgld RdSgvv.... sel..FDe.R nDAvkallsm
STAMA PPSA       .......snd ltqL.iLgad RdSnil.... aemgyFDe.R dpAvlagikm
METJA PPSA       .......lyr lgiiprLryk RsStatiyfn nnletilq.R trrikldklk
Consensus        ---------- ---L--L--- R-S------- -----FD--R -DA-------
signature #2 continued:SND LXQL SLGXX R
                        T   I  I TIA
                            V  V AVS
                            M  M LM
                               F IF
                               Y V
                               G
                               T

                851                                                       900
SMA ORF3         nlhaeeppma iVGpy..tPk Vkaptqddkt irs.ltgLaa sPgnVVakaR
STRCO PEPUE      nkrlpepplt fVGvhditaS Vr.paadg.. .ar.legLaa sPgriVgraR
ECOLI PPSA       airaakkqgk yVGIcgqgPS dhe.dfaawL meegidsLsl nPdtVVqtwl
STAMA PPSA       iiekahskga tVsIcgqaPS Vyp.eivefL veagidsisv nPdaViatrR
METJA PPSA       efkkpvedkk lidIsqilPe lkefdykgyL yktykekLfi g...inklee
Consensus        ---------- -VGI----PS V--------L -------L-- -P--VV---R

                901                                                       950
SMA ORF3         vitdLqvqag efqpdeIl.v arftDaswtp lfalaagivt digstlshsc
STRCO PEPUE      ivedLvwqad efeageIl.v tgytDaswtp lfaiaggvvt digsmlshss
ECOLI PPSA       slaeLkk--- ---------- ---------- ---------- ----------
STAMA PPSA       lvasierkim lkrlnkImdk lnklelgf-- ---------- ----------
METJA PPSA       ylskidkd.g ierikqkikl lkesDiysir ikkvgedyge vynitvkaen
Consensus        ----L----- ------I--- ----D----- ---------- ----------

                951                                                      1000
SMA ORF3         ivarefgipa vvnlktatqi insgdmlild gdsgtviiqh qeernhdg--
STRCO PEPUE      iva------- ---------- ---------- ---------- ----------
ECOLI PPSA       ---------- ---------- ---------- ---------- ----------
STAMA PPSA       ---------- ---------- ---------- ---------- ----------
METJA PPSA       efnhnyvvwt khytpivvfn chaaivsrel gtpcvvgtkk atkvlkdgmi
Consensus        ---------- ---------- -↑-------- ---------- ----------
                                       active site catalytic residue

                1001                                                     1050
SMA ORF3         ---------- ---------- ---------- ---------- ----------
STRCO PEPUE      ---------- ---------- ---------- ---------- ----------
ECOLI PPSA       ---------- ---------- ---------- ---------- ----------
STAMA PPSA       ---------- ---------- ---------- ---------- ----------
METJA PPSA       vtvdgekgiv yegeikkvee kekkqevvvq qapiitatev kvnvsmpeva
Consensus        ---------- ---------- ---------- ---------- ----------
```

(Figure 4.8a continued...)

```
                1051                                                1100
SMA ORF3        ---------- ---------- ---------- ---------- ----------
STRCO PEPUE     ---------- ---------- ---------- ---------- ----------
ECOLI PPSA      ---------- ---------- ---------- ---------- ----------
STAMA PPSA      ---------- ---------- ---------- ---------- ----------
METJA PPSA      eraaatgadg vgllraehmi lglgkhprki leeegeeali ealmegirkv
Consensus       ---------- ---------- ---------- ---------- ----------

                1101                                                1150
SMA ORF3        ---------- ---------- ---------- ---------- ----------
STRCO PEPUE     ---------- ---------- ---------- ---------- ----------
ECOLI PPSA      ---------- ---------- ---------- ---------- ----------
STAMA PPSA      ---------- ---------- ---------- ---------- ----------
METJA PPSA      adafyprpvt yrtldaptde frgleggene piehnpmlgw rgirrdldev
Consensus       ---------- ---------- ---------- ---------- ----------

                1151                                                1200
SMA ORF3        ---------- ---------- ---------- ---------- ----------
STRCO PEPUE     ---------- ---------- ---------- ---------- ----------
ECOLI PPSA      ---------- ---------- ---------- ---------- ----------
STAMA PPSA      ---------- ---------- ---------- ---------- ----------
METJA PPSA      dilkcelkai krlreegykn ieimiplvth pdevrrvkei mrevglepck
Consensus       ---------- ---------- ---------- ---------- ----------

                1201                                                1250
SMA ORF3        ---------- ---------- ---------- ---------- ----------
STRCO PEPUE     ---------- ---------- ---------- ---------- ----------
ECOLI PPSA      ---------- ---------- ---------- ---------- ----------
STAMA PPSA      ---------- ---------- ---------- ---------- ----------
METJA PPSA      dipfgimvet paaaliiedf ikeginfvsl gtndltqyti aidrnnelvs
Consensus       ---------- ---------- ---------- ---------- ----------
                           signature #2: DXLSL GSNDLXQLSL GXXR
                                         E I I  T   I ITI A
                                         Q V V      V VAV S
                                         S M M      M MLM
                                           F F        FIF
                                                      YV
                                                      G
                                                      T

                1251                                                1300
SMA ORF3        ---------- ---------- ---------- ---------- ----------
STRCO PEPUE     ---------- ---------- ---------- ---------- ----------
ECOLI PPSA      ---------- ---------- ---------- ---------- ----------
STAMA PPSA      ---------- ---------- ---------- ---------- ----------
METJA PPSA      kyykedhpav lklvehvikt ckkhgiktsi cgqagsrphi veklvewgid
Consensus       ---------- ---------- ---------- ---------- ----------

                1301                                1336
SMA ORF3        ---------- ---------- ---------- ------
STRCO PEPUE     ---------- ---------- ---------- ------
ECOLI PPSA      ---------- ---------- ---------- ------
STAMA PPSA      ---------- ---------- ---------- ------
METJA PPSA      svsanidave tirrvvarte qkvilnyirk syvere
Consensus       ---------- ---------- ---------- ------
```

Figure 4.8b

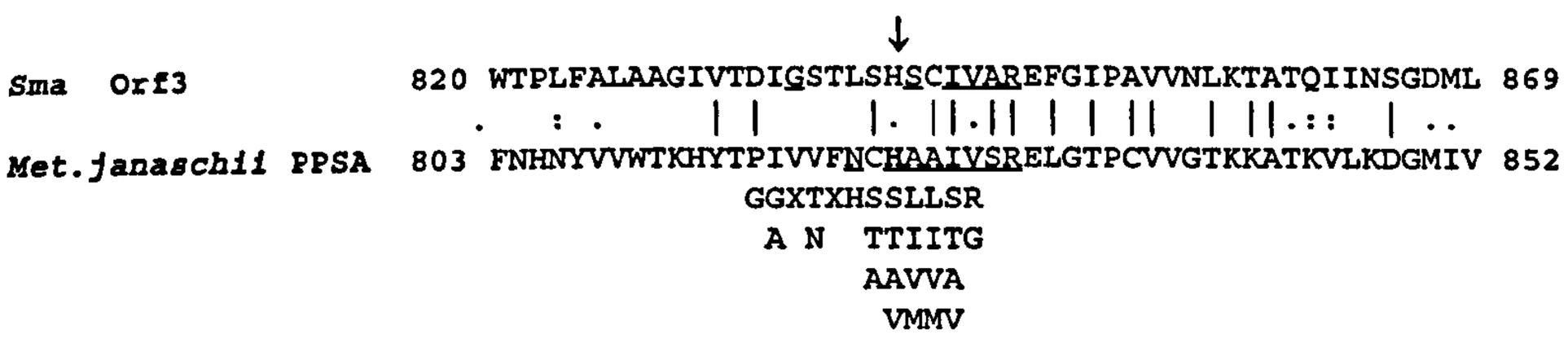


Figure 4.8b Region of an alignment of Orf3 and *Methanococcus janaschii* PEP synthase.

For clarity, only the aligned region around the catalytic His^{824} (in **bold type**) of *Met. janaschii* PEP synthase (PPSA) is shown. A histidine residue in the Orf3 sequence marked by an arrow lies at the same relative position in this stringent alignment and the region around it bears some resemblance to the signature sequence. Underlined residues in both sequences represent identity with the signature sequence. See text in section 4.6.3 for details.

[LIVMFYGT]-[STALIV]-[LIVMF]-[GAS]-x(2)-R (PROSITE Documents PDOC00527 and PS00370).

In the multiple alignment presented in Fig. 4.8.2a, it can be seen that there is a good similarity between the proteins in the N-terminal region, although not at the extreme N-termini. It is also obvious throughout the alignment that the *Str. coelicolor* A3(2) and *S. marcescens* proteins also differ from the other homologues, and importantly, many differences are seen around the signature areas. Since the first of these signature areas surrounds the proposed catalytic region, there are important potential implications in these findings. The first signature was easily located in the *E. coli* and *Sta. marinus* amino acid sequences due to their known catalytic histidine residues: His ^{447} in *Sta. marinus* (Cicicopol *et al.*, 1994) and His^{421} in *E. coli* (Niersbach *et al.*, 1992). It can be seen that the amino acid sequences of these two proteins conform to the signature. However, the Orf3 sequence (and that of PEPUE of *Str. coelicolor* A3(2)) differs considerably, even lacking a histidine residue at the same relative position (see Fig 4.8a). The designated catalytic histidine residue of *Met. janaschii*, is located closer to the C-terminus of the protein (His^{824}, Bult *et al.*, 1996), so consequently, the second signature is also found still further down in this protein, so these areas do not align in the multiple alignment with the other proteins in Fig. 4.8a. It can also be seen that residues Val^{822} and Val^{823} in the *Met. janaschii* sequence do not correspond to the signature sequence (G-[GA]-).

In the second signature, there are differences between the amino acid sequences of *S. marcescens* and *Str. coelicolor* A3(2) in comparison to those of *E. coli* and *Sta. marinus*. Firstly, at aligned position 794 in Fig. 4.8a, Orf3 has isoleucine at this relative position where the other homologues contain either aspartate or glutamate; similarly, PEPUE of *Str. coelicolor* A3(2) contains valine. Six residues to the right of this position in Fig 4.8a, it can be seen that Orf3 and PEPUE contain an addition nine residues not seen in the other homologues, so consequently this results in a complete departure from the consensus.

So far as the signature sequence around the catalytic histidine residue is concerned, it can be seen in Fig. 4.8a that Orf3 and PEPUE lack histidine residues at the same relative positions in comparison to *E. coli* and *Sta. marinus*.

However, it was found that when the Orf3 and PPSA (*Met. janaschii*) are aligned separately, as illustrated in Fig. 4.8b, a conserved histidine residue is seen at the same relative position in the alignment, but once again, deviations from the signature sequence are seen. The sequence surrounding every histidine residue in the Orf3 sequence was examined to see whether the putative catalytic domain might be found elsewhere but this did not prove fruitful and furthermore, no consensus for the second signature was found in the separate alignment in Fig.4.8b. Separately aligned sequences of Orf3 and PEPUE (data not shown) were examined for an alternative location for the second signature sequence, based on the assumption that the regions of these proteins aligned in Fig. 4.8a and Fig. 4.8b might be coincidentally similar to that of *E . coli* and *Sta. marinus*. Once again, this did not yield any alternatives.

The multiple alignment presented in Fig. 4.8a was generated with low stringency to overcome the dissimilarities between the sequences, so the information which can be derived from this is limited. It therefore appears that Orf3, while having regions in common with the PEP utilising enzymes in the database, also differs considerably from them and it is possible that these differences relate to its function as a dedicated prodigiosin biosynthetic protein in *S. marcescens*. This possibility appears more plausible in view of the fact that the PEPUE sequence of *Str. coelicolor* A3(2) also differs from the "conventional" PEP utilising enzymes in a similar manner in the same relative regions and the gene encoding this putative protein is located in the proximity of the *red* cluster in this organism.

4.7.1 THE *orf4* PREDICTED GENE PRODUCT

Amino acid translation of Orf4 showed that it encodes a putative protein which is 866 amino acids in length. It has a predicted MW of 96977.1Da and a theoretical pI of 5.66.

FIGURE 4.9 PREDICTED SECONDARY STRUCTURE OF Orf4

Shown opposite is the "squiggle plot" output from UWGCG PROTEINSTRUCTURE and PLOTSTRUCTURE. Regions of hydrophobicity are represented by diamonds and regions of hydrophilicity by ovals.
Key to symbols:

SINE WAVE : α helices

SHARP SAW-TOOTH : β-turns

DULL SAW-TOOTH : random coils

Turns are shown as a 180° turn in the plotting line.

PLOTSTRUCTURE of: crf4.pr ck: 2854

4.7.2 CELLULAR LOCATION AND PREDICTED SECONDARY STRUCTURE

The secondary structure of Orf4 is represented in Fig. 4.9. SIGNALP and PSORT did not reveal a cleavable N-terminal signal sequence and this is supported by the absence of hydrophobicity at the N-terminus, with the exception of the first few residues (Fig. 4.9). It is also predicted by PSORT analysis that there are no likely transmembrane regions in the protein so it is therefore predicted to be cytoplasmic.

4.7.3 DATABASE SEARCH RESULTS FOR Orf4

Screening of the non-redundant amino acid database in BLAST searches failed to find any significant homologies to Orf4. It therefore was concluded that Orf4 is a hypothetical protein of unknown function. However, other work done in this study (presented in Chapter 6), suggests that Orf4 is required for prodigiosin biosynthesis because a transposon-induced non-pigmented mutant was generated in which the insertion has been mapped to *orf4*. However, another possibilty is that since *orf4* appears to be translationally coupled to *orf5*, this transposon insertion would have a polar effect on the latter, and the phenotype may actually be due to this reason.

4.8.1 THE *orf5* PREDICTED GENE PRODUCT

The predicted protein encoded by *orf5* is 853 amino acids in length. It has a MW of 93577.3Da and a theoretical pI of 5.6.

4.8.2 PREDICTED CELLULAR LOCATION AND SECONDARY STRUCTURE

The predicted secondary structure is shown as a squiggle plot in Fig. 4.10. Analysis of the amino acid sequence by SIGNALP and PSORT predicts that there is no N-terminal signal sequence present, and the protein is predicted to be located in the cytoplasm. This is supported by the absence of hydrophobicity at the N-terminus (see Fig. 4.10), which was confirmed by PROTSCALE analysis (data not shown).

FIGURE 4.10 PREDICTED SECONDARY STRUCTURE OF ORF5

Shown opposite is the "squiggle plot" output from UWGCG PROTEINSTRUCTURE and PLOTSTRUCTURE. Regions of hydrophobicity are represented by diamonds and regions of hydrophilicity by ovals.
Key to symbols:

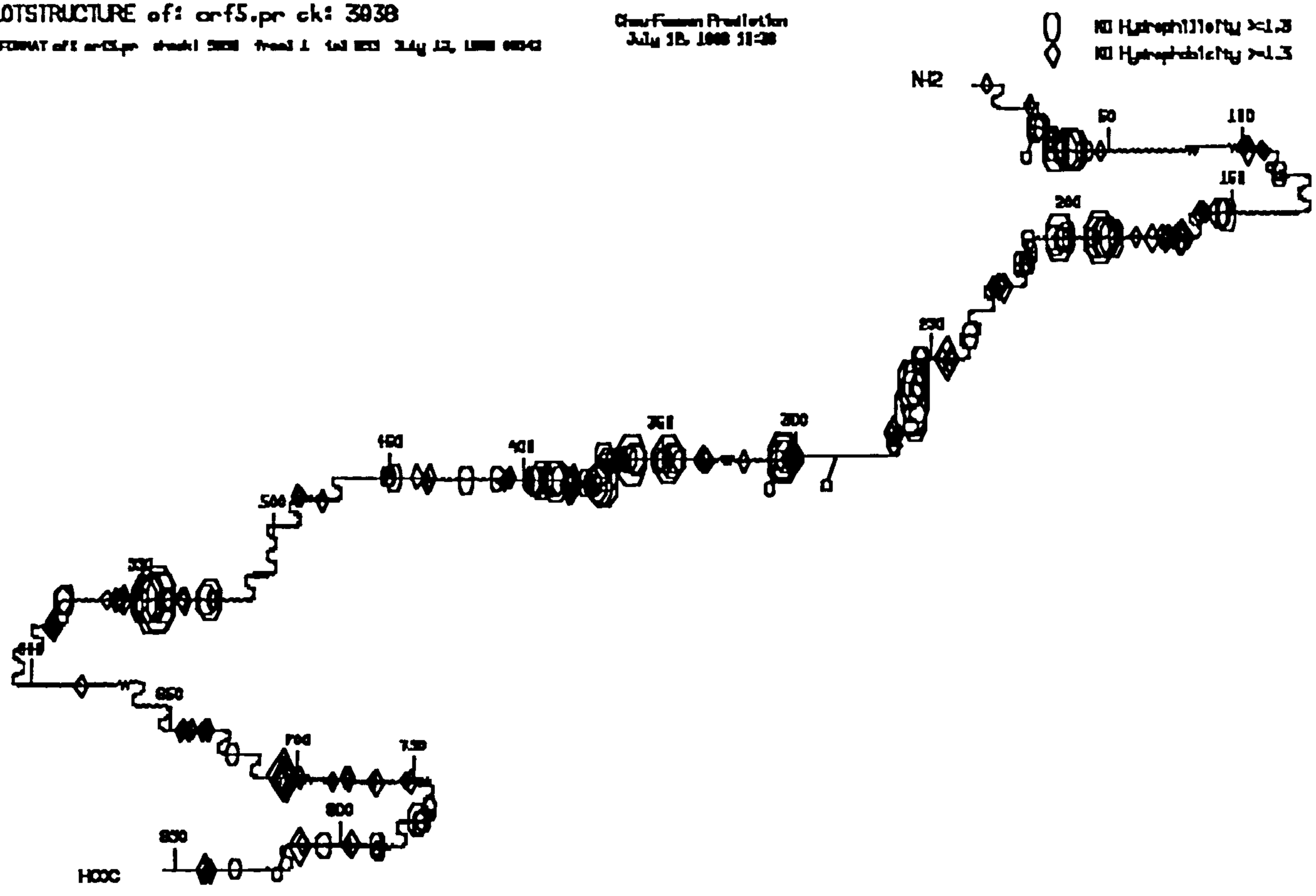
SINE WAVE : α helices

SHARP SAW-TOOTH : β-turns

DULL SAW-TOOTH : random coils

Turns are shown as a 180° turn in the plotting line.

PLOTSTRUCTURE of: crf5.pr ck: 3938



4.8.3 DATABASE HOMOLOGUES OF Orf5

Screening of the non-redundant amino acid database using BLAST revealed that Orf5 has significant homology to prokaryotic and eukaryotic Class III (pyridoxal phosphate dependent) aminotransferases. The greatest similarity is to probable ornithine amino transferase of *E. coli* (EC 2.6.1.13, Blattner *et al.*, 1997, direct database submission). Other high-scoring BLAST hits include ArgD (acetylornithine aminotransferase EC 2.6.1.11) from the thermophilic species *Met. janaschii* (Bult *et al.*, 1996), in which this enzyme is involved in the biosynthesis of arginine from glutamate (Bult *et al.*, 1996, Cunin *et al.*, 1986); RocD (ornithine aminotransferase EC 2.6.1.13) of *Bacillus subtilis*, which is involved in arginine catabolism (Gardan *et al.*, 1995; Calogero *et al.*, 1994), and ornithine aminotransferase precursor (EC 2.6.1.13) of mouse (Manenti *et al.*, 1992, direct database submission). These homologues and their details are summarised in Table 4.4. A multiple alignment of Orf5 and its database homologues from Table 4.4 is presented in Fig. 4.11

Pyridoxal phosphate dependent enzymes, including Class III of the aminotransferases, have this co-factor covalently bonded to a conserved lysine residue (Styer, 1988c). Examples of Class III aminotransferases include all those listed in Table 4.8. It is a conserved family in evolutionary terms, with good amino acid homology in both prokaryotes and eukaryotes. Ornithine aminotransferase (EC 2.6.1.13) catalyses the transfer of an amino group from L-ornithine to 2-oxo-acid, yielding glutamate-5-semialdehyde and an L-amino acid (Gardan *et al.*, 1995; Calogero *et al.*, 1994); similarly, acetylornithine aminotransferase (EC 2.6.1.11) catalyses the transfer of an amino group from N2-acetyl-L-ornithine to 2-oxo-glutarate, yielding N-acetyl-L-glutamate-5-semialdehyde and L-glutamate.

It can be seen in Fig. 4.11 that there is homology throughout the alignment between the different homologues. However, Orf5 has a much longer predicted N-terminal region, comprising ~310 additional residues in comparison with OAT from *E. coli*, for example. It is therefore possible that this protein has other function(s) in addition to those it is likely to share with the database homologues, and this could well be related to its role in prodigiosin biosynthesis.

TABLE 4.4 HOMOLOGUES OF ORF5 AS IDENTIFIED IN A BLAST SEARCH AGAINST THE NON-REDUNDANT AMINO ACID DATABASE

HOMOLOGUE	P-SCORE IN BLAST OUTPUT	% IDENTITY NUCLEIC ACID	AMINO ACID (OVERLAP) AND % IDENTITY SIMILARITY	DATABASE, ACCESSION NUMBER AND REFERENCE
probable ornithine amino transferase (OAT) *Escherichia coli*	$6.2x10^{-74}$	48.5	(496) 33.5 44.4	SWISSPROT P42588 Blattner *et al.*, 1997
acetyl-ornithine amino transferase (ArgD) *Methaococcus janaschii*	$2.2x10^{-71}$	44.5	(395) 33.6 45.6	SWISS-PROT Q58131 Bult *et al.*, 1996
ornithine amino transferase (RocD) *Bacillus subtilis*	$1.6x10^{-64}$	46.6	(399) 32.1 42.6	SWISS-PROT P38021 Gardan *et al.*, 1995 Calogero *et al.*, 1994
ornithine aminotransferase precursor (OATP) *Mouse*	$9.1x10^{-50}$	37.1	(439) 26.6 38.8	SWISS-PROT P29758 Manenti *et al.*, 1992, direct database submission

FIGURE 4.11 MULTIPLE ALIGNMENT AND CALCULATED CONSENSUS OF ORF5 AND ITS HOMOLOGUES

This represents the PRETTY output for Orf5 aligned with the bacterial homologues listed in Table 4.7.

Key: *BACSU* ROCD: *Bacillus subtilis* ornithine aminotransferase; *MOUSE* OAT: mouse ornithine aminotransferase; *METJA* ARGD: *Methanococcus janaschii* acetylornithine amino transferase; *ECOLI* OAT: *Escherichia coli* ornithine aminotransferase; *SMA* ORF5: *Serratia marcescens* Orf5

The proposed pyridoxal-phosphate attachment site is shown in **bold type** and is marked by an arrow. Also shown in **bold type** are the two conserved signature regions which identify this family of enzymes. The proposed active site residue critical for catalysis is marked by an arrow and the different residues which may be present in the conserved regions are shown below the relevant positions above or below the. See text in section 4.6.3 for details.

Figure 4.11

```
                    1                                                  50
BACSU ROCD  ---------- ---------- ---------- ---------- ----------
MOUSE OAT   ---------- ---------- ---------- ---------- ----------
METJA ARGD  ---------- ---------- ---------- ---------- ----------
ECOLI OAT   ---------- ---------- ---------- ---------- ----------
SMA ORF5    mkfgfiahpt svglkryvkm idllqrnste lhsgykrdlw rrenlvpfmn
Consensus   ---------- ---------- ---------- ---------- ----------

                   51                                             100
BACSU ROCD  ---------- ---------- ---------- ---------- ----------
MOUSE OAT   ---------- ---------- ---------- ---------- ----------
METJA ARGD  ---------- ---------- ---------- ---------- ----------
ECOLI OAT   ---------- ---------- ---------- ---------- ----------
SMA ORF5    fakitsatga tcegvikymp lvademlada rgianrvvsg ieelvedgae
Consensus   ---------- ---------- ---------- ---------- ----------

                  101                                             150
BACSU ROCD  ---------- ---------- ---------- ---------- ----------
MOUSE OAT   ---------- ---------- ---------- ---------- ----------
METJA ARGD  ---------- ---------- ---------- ---------- ----------
ECOLI OAT   ---------- ---------- ---------- ---------- ----------
SMA ORF5    lvglggftsi vgrrgeatae kspvpvtsgn slttyagyka lmqiqswldi
Consensus   ---------- ---------- ---------- ---------- ----------

                  151                                             200
BACSU ROCD  ---------- ---------- ---------- ---------- ----------
MOUSE OAT   ---------- ---------- ---------- ---------- ----------
METJA ARGD  ---------- ---------- ---------- ---------- ----------
ECOLI OAT   ---------- ---------- ---------- ---------- ----------
SMA ORF5    qpeqepvaiv gypgsiclal srlllaqgfs lhllhraghk dedellshlp
Consensus   ---------- ---------- ---------- ---------- ----------

                  201                                             250
BACSU ROCD  ---------- ---------- ---------- ---------- ----------
MOUSE OAT   ---------- ---------- ---------- ---------- ----------
METJA ARGD  ---------- ---------- ---------- ---------- ----------
ECOLI OAT   ---------- ---------- ---------- ---------- ----------
SMA ORF5    eqyrsrvtlt sdpedlyprc klfvaatsag gvidpyklqp gsvfidvalp
Consensus   ---------- ---------- ---------- ---------- ----------

                  251                                             300
BACSU ROCD  ---------- ---------- ---------- ---------- ----------
MOUSE OAT   ---------- ---------- ---------- ---------- ----------
METJA ARGD  ---------- ---------- ---------- ---------- ----------
ECOLI OAT   ---------- ---------- ---------- ---------- ----------
SMA ORF5    rdinsdtrpd rddiliidgg cvtatdavkl ggeslnvtik qqlngcmaet
Consensus   ---------- ---------- ---------- ---------- ----------

                  301                                             350
BACSU ROCD  ---------- ---------- ---------- ---------- ----------
MOUSE OAT   ---------- ---------- ---------- ---------- ----------
METJA ARGD  ---------- ---------- ---------- ---------- ----------
ECOLI OAT   ---------- --mitefvfi pifaiaagva qslqylnryh vireppehil
SMA ORF5    ivlalenrre nfslgrylal dnvleigela ekhgflvypl asygeridrq
Consensus   ---------- ---------- ---------- ---------- ----------
```

(Figure 4.11 continued...)

```
                351                                                 400
BACSU ROCD      ---------- ---------- ---------- ---------- mtalskskei
MOUSE OAT       -------mls klaslqtiaa lrrgvhtsva satsvatkkt eqgppsseyi
METJA ARGD      ---------- ---------- ---------- ---------- ----msqenw
ECOLI OAT       nrlpssasal acsahalnli ekrtldheem kalnreviey fkehvnpgfl
SMA ORF5        rvinlkryyh hdiysdepdt eqppasqlaf idaiiaqdpa redtldryhq
Consensus       ---------- ---------- ---------- ---------- ----------

                401                                                 450
BACSU ROCD      idqtShYgan nYhpLPiVis ealGawvkDp eGneYmDmLs aYsavNqGHr
MOUSE OAT       fereSkYgah nYhpLPVale rGkGiymwDv eGrqYfDfLs aYGavsqGHc
METJA ARGD      idlekkYhlq iYgrLPVVlv eGkGmevyDi dGkkYlDfLa GiGvnNvGHc
ECOLI OAT       eyrkSvtagg dYg..aVewq aGslntlvDt qGqefiDcLg GfGifNvGHr
SMA ORF5        finpmmvefl klqhcdnVfr rasGtqlfta dGeaflDmva GYGciNlGHn
Consensus       ----S-Y--- -Y--LPVV-- -G-G----D- -G--Y-D-L- GYG--N-GH-

                451                                                 500
BACSU ROCD      HPKIIqAlKd QaDKitLTSr afhn..dqlg pfyEktakLt ..gkemilpm
MOUSE OAT       HPKIIdAmKs QvDKltLTSr afyn..nvlg eyeEyitkLf ..nynkvlpm
METJA ARGD      HPKvveAiKk QaetlihTSn iyyt..ipqi kLAkkLveLs ..gldraFFc
ECOLI OAT       nPvvvsAvqn QlaKqpLhSq elld..plra mLAktLaaLt pgklkysFFc
SMA ORF5        pqpIIdAlKa ylDaqgpnfi qyisipeqaa kLAEvLchfa pgnmgrvFFs
Consensus       HPKII-A-K- Q-DK--LTS- ---------- -LAE-L--L- -------FF-

                501                                                 550
BACSU ROCD      NtGaEAVEsA vKaARrwaYe vKgvadnqAe IIacvGnFHG RTmlAvSlss
MOUSE OAT       NtGvEAgEtA cKLARrwgYt vKgiqkyKAk IvfadGnFwG RTLsAiSsst
METJA ARGD      NSGaEAnEgA iKfARk..Yv sKvlgregge IIsmynaFHG RTLttLaaTp
ECOLI OAT       NSGtEsVEaA lKLAkayqsp .....rgKft fIatsGaFHG ksLgALSaTa
SMA ORF5        NSGtEAVEaA mKLAka.... ....stgKAg IaylknsyHG kTLgALSiTg
Consensus       NSG-EAVE-A -KLAR---Y- -K-----KA- II---G-FHG RTL-ALS-T-

                551                                                 600
BACSU ROCD      eeeykrGFgP mLPGikliPy gDvEALRqAi .t.....pnt AAflfEPIQG
MOUSE OAT       dptsydGFgP fmPGFetiPy nDlpALerAL qd.....pnv AAfMvEPIQG
METJA ARGD      kpkyqdGFyP LpPGFkyVPF nDIEALkeAi td......kt AAiMiEPvQG
ECOLI OAT       kstfrkpFmP LLPGFrhVPF gnIEAmRtAL neckktgddv AAvilEPIQG
SMA ORF5        rekhrrhFkP LLasmieVPF aDIEALRqtL s.....rddi gAlMiEPIQG
Consensus       ------GF-P LLPGF--VPF -DIEALR-AL ---------- AA-M-EPIQG

                601                                                 650
BACSU ROCD      EaGiviPPeG fLqeaaaiCk eenvLfIaDE iQTGLGRTGK tFACdwdGiv
MOUSE OAT       EaGViVPdPG YLtgVReLCt rhqvLfIaDE iQTGLaRTGr wlAvdHEnVr
METJA ARGD      EGGihVadkd YLkaVRdLCd dknivlIfDE VQcGmGRTGr mFAfEHyGVe
ECOLI OAT       EGGVilPPPG YLtaVRkLCd efgaLmIlDE VQTGmGRTGK mFACEHEnVq
SMA ORF5        EGGVhVPPPG YLrtVqeiCr qtdtLlmvDE VQTGLGRTGK lFACEwEGie
Consensus       EGGV-VPPPG YL--VR-LC- ----L-I-DE VQTGLGRTGK -FACEHEGV-

SIGNATURE SEQUENCE: LLXDE LXXGLXSXGX XXXXXXXXXX
                    II    I  PI A S
                    VV    V   V C A
                    MM    M   M R D
                    FF    A   F
                    YY        W
                    WW        A
                    CC        G
```

(Figure 4.11 continued...)

```
                          pyridoxal phosphate attachment site
                    651       ↓                                                  700
      BACSU ROCD    PDmyiLgKAL GGGVfPIsci aadrEIlg.. .vf..nPGsH GSTFGGNPLA
      MOUSE OAT     PDmvlLgKAL sGGlyPvsAv lcddEIml.. .ti..kPGeH GSTyGGNPLg
      METJA ARGD    PDiLtLaKAL GGGV.PIGAv vlkeEIak.. .al..syGdH GtTFGGNPLA
      ECOLI OAT     PDiLcLaKAL GGGVmPIGAt iateEvfs.. .vlfdnPflH ttTFGGNPLA
      SMA ORF5      PDvLmLsKsL sGGVmPIGAt lcraifgngp ygtadrflmH sSTFGGgniA
      Consensus     PD-L-L-KAL GGGV-PIGA- ----EI---- ------PG-H GSTFGGNPLA
(continued...)      XDLXXXGKXX XGG
                      I   S     SS
                      V   A     TA
                      M         A
                      F         D
                      Y         N
                      W
                      C

                    701                                                          750
BACSU ROCD    CavsiAsLEV LEdekLadrs lelGeYFkse Lesi..ds.p vikEVRGrGL
MOUSE OAT     CriAiAaLEV LEEenLaenA dkmGailrke LmkL..ps.d vVtsVRGkGL
METJA ARGD    CsaALAsvEV iEElikddkv iekGkYFirk LenLiekY.n fikEVRGlGL
ECOLI OAT     CaaALAtinV LlEqnLpaqA eqkGdmlldg frqLareYpd lVqEaRGkGm
SMA ORF5      avvALsaLre ilaqdLvgnA erlGtYFkqa LtdvaarYp. fVaEiaGrGL
Consensus     C--ALA-LEV LEE--L---A ---G-YF--- L--L---Y-- -V-EVRG-GL

                    751                                                          800
BACSU ROCD    fiGvEltea. ...Arpycer lkEeGlLcke ThdTvIRFaP PLiiskEdld
MOUSE OAT     lnaIviretk dcdAwkvclr lrdnGlLakp ThgdiIRlaP PLvikedeir
METJA ARGD    miGaEl.... efngadivkk mlEkGfLinc TsdTvlRFlP PLivekEhid
ECOLI OAT     lmaIEfvdne igynfasemf rqrvlvagtl nnaktIRieP PLtltiEqce
SMA ORF5      mlGIqfdqt. ..fAgavgas arEfatrlpg dwhTtwkFlP dpv..qahlk
Consensus     --GIE----- ---A------ --E-G-L--- T--T-IRF-P PL----E---

                    801                                                          850
BACSU ROCD    waIEkikhvL rna~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
MOUSE OAT     esvEiinkti lsf~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
METJA ARGD    alInaldevf teikk~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
ECOLI OAT     lvIkaarkaL aamrvsveea ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
SMA ORF5      aamErmeqsL gemfcmkfvt klcqdhnilt fitansstvi riqppltisk
Consensus     --IE-----L ---------- ---------- ---------- ----------

                    851                       872
BACSU ROCD    ~~~~~~~~~~ ~~~~~~~~~~ ~~
MOUSE OAT     ~~~~~~~~~~ ~~~~~~~~~~ ~~
METJA ARGD    ~~~~~~~~~~ ~~~~~~~~~~ ~~
ECOLI OAT     ~~~~~~~~~~ ~~~~~~~~~~ ~~
SMA ORF5      aeidrfvsaf atvcdelstf le
Consensus     ---------- ---------- --
```

As with Orf2, the alternative possibility is that Orf5 might in fact represent two proteins, but on the basis of available sequence data, it is not possible to resolve this. It is therefore assumed that Orf5 represents one protein. It can also be seen in Fig. 4.8.2 that Orf5 possesses ~50 additional residues at its C-terminus which may also be related to its biological role.

It has been found by similarity searches that the sequence around the pyridoxal-phosphate attachment site is sufficiently conserved in this enzyme family to deduce a signature sequence: [LIVMFYWC](2)-D-E-[LIVMA]-x(2)-[GP]-x(0,1)-[LIVMFYWAG]-x(0,1)-[SACR]-x-[GSAD]-x(12,16)-D-[LIVMFYWC]-x(2,3)-[GSA]-K-[GSTADN]-[GSA] (PROSITE Documents PDOC00519, PS00600). A lysine residue comprising the pyridoxal-phosphate attachment has been identified in the database proteins, in many cases by similarity: in *E. coli* OAT, it is Lys^{270} (Blattner *et al.*, 1997, direct database submission); in *Met. janaschii* ARGD it is proposed to be Lys^{252} (Bult *et al.*, 1996); in *B. subtilis* it is Lys^{258}(Gardan *et al.*, 1995; Calogero *et al.*, 1994) and in mouse OATP it is Lys^{292} (database accession information). The relevant lysine residues and the surrounding signature sequence are shown in Fig. 4.11. In Fig. 4.11, it can be seen that the region around the proposed pyridoxal-phosphate attachment site in the database proteins is highly conserved, and upon alignment, the same conservation is seen in *S. marcescens* Orf5. It is highly probable on the evidence of this data therefore that Orf5 is an aminotransferase, albeit possibly a modified one, if it does indeed possess additional residues at its N- and C-termini.

4.9 DISCUSSION OF CHRACTERISTICS OF PUTATIVE PROTEINS BY THE *Pig* CLUSTER

In order to discuss these findings in the wider context of prodigiosin biosynthesis, it is necessary to mention the putative proteins were identified in the 3' end of the *pig* cluster by A. Cox. In the latter half of the cluster, it was found that many more putative ORFs are present (see Fig. 3.7), generally encoding smaller proteins than those identified in this study (A. Cox. Pers. comm). A representation of the predicted proteins and their closest database homologues is given in Fig. 4.12. Table 4.5 contains summary information regarding the

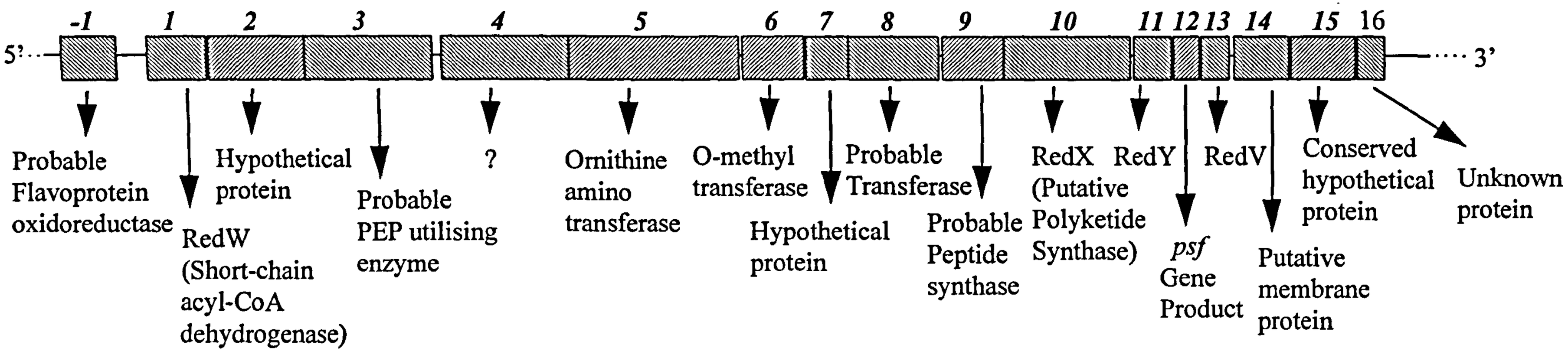


Figure 4. 12 HOMOLOGUES OF THE PREDICTED PROTEINS OF THE *Serratia marcescens pig* CLUSTER

ORFs are represented by the shaded boxes and are numbered accordingly. Also shown is *orf(-1)*, upstream of the cluster. The database proteins to which significant homologies were found are arrowed where applicable, otherwise (?) is shown. See text in section 4.9 for details.

TABLE 4.5 SUMMARY INFORMATION OF THE PUTATIVE ORFs OF THE 3' END OF THE *pig* CLUSTER

Information was derived from searches of the non-redundant amino acid database using the translated amino acid sequences of the predicted coding regions. See also text in section 4.9 for further details.

Table 4.5

PUTATIVE *pig* CLUSTER PROTEIN	HIGHEST SCORING HOMOLOGUE, (BLAST P-SCORE) AND SPECIES	COMMENTS	DATABASE, ACCESSION, AND REFERENCE(S)
Orf6	Tetracenomycin polyketide synthesis 8-*O*-methyl-transferase (TCMO), ($4.2x10^{-27}$) *Streptomyces glaucescens*	Catalyses the second step reaction in the synthesis of tetracenomycin C (TCMC) by converting TCMB3 to TCME. Also methylates TCMD3 into 8-O-methyl-TCMD3 in a side reaction.	SWISSPROT P39896 Summers *et al.*, 1992
Orf7	Hypothetical Protein SC3F7.09 ($1.7x10^{-13}$) *Streptomyces coelicolor* A3(2)	Identified in the *Str. coelicolor* A3(2) genome sequencing project; maps to the *red* locus and lies adjacent to SC3F7.08, encoding a probable acyl-carrier protein synthase and SC3F7.10 encoding a probable transferase (see Orf8 entry below).	EMBL AL021409 Redenbach *et al.*, 1996; Parkhill *et al.*, Oliver and Harris, direct database submission.
Orf8	SC3F7.10 Probable transferase ($4.5x10^{-199}$) *Streptomyces coelicolor* A3(2)	See Orf7 entry above. The downstream gene in *Str. coelicolor* A3(2) encodes SC3F7.11, a probable peptide synthase (sse Orf9 enrty below).	EMBL AL021409 References as per Orf7 entry.
Orf9	SC3F7.11 Probable Peptide Synthase ($3.6x10^{-15}$) *Streptomyces coelicolor* A3(2)	See Orf8 entry above. The downstream gene in *Str. coelicolor* A3(2) SC3F712 encodes a probable polyketide synthase.	EMBL AL021409 References as per Orf7 entry
Orf10	RedX (SC2E9.19) ($2.6x10^{-76}$) *Streptomyces coelicolor* A3(2)	Probable polyketide synthase; maps to the *red* locus in *Str. coelicolor* A3(2); ORF lies adjacent to *redW* and *redD*.	EMBL AL021530 References as per Orf7 entry

(Table 4.5 continued...)

PUTATIVE *pig* CLUSTER PROTEIN	HIGHEST SCORING HOMOLOGUE, (BLAST P-SCORE) AND SPECIES	COMMENTS	DATABASE, ACCESSION, AND REFERENCE(S)
Orf11	RedY (SC2E9.21)(4.4x10-27) *Streptomyces coelicolor* A3(2)	This is the only significant database homologue; maps to the *red* locus in *Str. coelicor* A3(2); ORF lies adjacent to *redW* and *redZ.*	EMBL AL021530 References as per Orf7 entry
Orf12	*psf* gene product ($8.3x10^{-5}$) *Bacillus pumilis*	Weak homology (79 amino acid overlap) to this protein; required for the production of surfactin antibiotic in *B. pumilis* (possible regulator). Belongs to the GSP/HETI/SFP family.	SWISSPROT P55810 Morikawa *et al.*, 1992
Orf13	RedV (SC3F7.02c) ($5.9x10^{-29}$) *Streptomyces coelicolor* A3(2)	Maps to the *red* locus of *Str. coelicolor* A3(2); ORF lies adjacent to *redZ* and a hypothetical protein gene encoding SC3F7.02c.	EMBL AL021409 References as per Orf7 entry
Orf14	SC310A53 Putative Membrane Protein ($1.1x10-^{23}$) *Streptomyces coelicolor* A3(2)	Maps to region close to the *red* locus in *Str. coelicolor* A3(2), adjacent to an ORF encoding probable PEP utilising enzyme (SC10A5.01) and an ORF encoding a putative membrane protein (SC10A5.03).	EMBL AL01529 References as per Orf7 entry
Orf15	AE001075.19 Conserved Hypothetical Protein ($8.5x10^{-35}$) *Archaeoglobus fulgidus*	Identified during the *A. fulgidus* genome sequencing project.	GENBANK AE00782 Klenk *et al.*, 1997
Orf16	AE000867.10 Unknown Protein($5.6x10^{-5}$) *Methanobacterium thermoautotrophicum*	Identified during the *Mba. thermoautotrophicum* genome sequencing project.	GENBANK AE000666 Smith *et al.*, 1997

predicted proteins of the 3' end of the *pig* cluster, and their homologues. Each of the putative ORFs identified in the *pig* cluster shall be considered in turn in the context of pigment biosynthesis. However, this discussion is purely speculative, based on the findings of the work presented in this chapter and in Chapter 3.

If Orf1 is an acyl-CoA dehydrogenase, its role in prodigiosin biosynthesis is not immediately obvious. One possibility is that it is involved in the early part of the biosynthetic pathway which involves the modification of central metabolic intermediates, and these might come from fatty acid metabolism or metabolism of branched-chain amino acids (see section 4.4.3). Another possibility is that it is involved in condensation reaction catalysis to link acyl moieties in the biosynthesis of prodigiosin, in a manner similar to that of condensing enzymes of polyketide biosynthetic pathways of other secondary metabolite producing species.

The possible role of Orf2, which is homologous to a hypothetical protein (SC3F7.05c) of unknown function from *Str. coelicolor* A3(2) cannot be speculated at present. In *Str. coelicolor* A3(2), the ORF encoding SC3f7.05c lies adjacent to an ORF which encodes a probable acyl-carrier protein and these genes are found close to known *red* genes of the undecylprodigiosin biosynthetic locus on the chromosome (Redenbach *et al.*, 1996; database accession EMBL AL02401)

If Orf3 is a PEP-utilising enzyme, it may be involved in an early step of the prodigiosin biosynthetic pathway, perhaps at the point where central metabolic intermediates are diverted into the pathways through which they are ultimately incorporated into prodigiosin. Furthermore, since strictly conserved protein domains are highly suggestive of biological function, the lack of identity with its homologues in otherwise conserved regions characteristic of PEP-utilising enzymes (see section 4.6.3) suggests that it is not simply a probable PEP-utilising enzyme of central metabolism and it is likely to function in a highly specialised way.

Orf4 does not have any database homologues at present. Experiments done in this study (presented in subsequent chapters) provide evidence that the *orf4* gene product is required for prodigiosin biosynthesis. Therefore, appears that Orf4 possibly is a protein which is unique to prodigiosin biosynthesis.

Orf5 is probably a modified ornithine amino-transferase. A possible role for such an enzyme in prodigiosin biosynthesis is more obvious because prodigiosin is known to be synthesised from various amino acids (see section 1.3.3). It may therefore be involved in the modification of amino acid precursors which are then incorporated into the pyrrole components of the molecule.

It cannot be doubted that Orf6, encoding a protein homologous to tetracenomycin polyketide synthesis O-methyltransferase is likely to be a specialised and dedicated enzyme of the biosynthetic pathway. One possible - and immediately obvious - role for it might be the transfer of a methyl group to replace hydroxyl side chain attached to position 4 of HBC, the only identified precursor to MBC, the bipyrrolic component of prodigiosin (see section 1.3.3). Previously characterised pigment mutants of *S. marcescens* which were blocked in this methylation step were found to produce instead the orange pigment norprodigiosin (Morrison, 1966, see also section 1.3.3). Future work which could be done in order to test this hypothesis could include in-frame mutagenesis of *orf6*, and assessing whether the resultant mutants have the predicted orange phenotype. It is also possible that Orf6 might be involved in side-chain methylation at another stage of pigment biosynthesis.

The only database homologue for Orf7 is a *Str. coelicolor* A3(2) hypothetical protein of unknown function, which maps to a chromosomal region which is proximal to known *red* (undecylprodigiosin-encoding) genes, therefore, no conclusions can be drawn at present about the possible role of Orf7 in pigment biosynthesis.

Orf8 has similarity to many database transferases, the greatest similarity being to a *Str. coelicolor* A3(2) probable transferase. As stated in Table 4.9, in *Str. coelicolor* A3(2) the ORF encoding this protein lies in the region adjacent to the *red* cluster; in this system, its role might be functionally related to that of the probable acyl carrier protein synthase encoded immediately upstream in this system. Homologues of both acyl carrier protein synthases and transferases are found in polyketide producing species and the encoding genes are clustered in the chromosome (Bibb *et al.*, 1989, see also sections 1.4.4 and 1.4.5); in these systems, the transferase loads the acyl and carboxy-acyl intermediates onto a

dedicated acyl carrier protein which is distinct from those of fatty acid metabolism and together they form the *β*-ketoacyl intermediates of the pathway (Bibb *et. al.*, 1989). However, Orf8 was *not* found to be homologous to polyketide biosynthesis transferases *per se*, nor was an acyl-carrier protein-encoding homologue found in the *pig* cluster. As another example, Orf8 was found to be homologous to 8-amino-7-oxononanoate synthetase (EC 2.3.1.47) (7-KAP synthetase), a *Bacillus sphaericus* enzyme encoded by *bioF* which catalyses an intermediate step in the biosynthesis of biotin by the addition of 6-carboxy-hexanoyl-CoA to alanine to form 8-amino-7-oxononanoate (Gloecker *et al.*, 1990). Similarly, in *S. marcescens*, such an enzyme might be involved in the transfer of acyl groups during prodigiosin biosynthesis.

Orf9 is homologous to a probable peptide synthase of *Str. coelicolor* A3(2) , encoded by an ORF which, once again, is located proximal to known *red* genes on the *Str. coelicolor* A3(2) chromosome. Peptide synthases are widespread in secondary metabolite producing organisms, particularly those producing peptide antibiotics. Other homologues of Orf9 include tyrocidin synthetase of *Brevibacillus brevis* (Mootz and Mahariel, 1997); PPS1, a peptide synthase of *Bacillus subtilis* (Tognoni *et al.*, 1995); GRSA, (gramicidin-S) synthase of *Bre. brevis* which activates phenylalanine and racaemises it to the *D*-isomer form in the first step of antibiotic biosynthesis (Kraetzschmar *et al.*, 1989). The common features shared by these enzymes is that they are large, multi-functional proteins which use phosphopantetheine as a co-factor; they are also involved in activation of amino acids. It is possible that Orf9, which is predicted to be a large protein, might also act in a multi-functional way in prodigiosin biosynthesis.

Orf10 is a homologue of RedX, one of the proteins involved in undecylprodigiosin biosynthesis in *Str.coelicolor* A3(2). This is a putative polyketide synthase, by similarity (Table 4.5). Polyketide synthases are large multimeric multifunctional enzymes involved in the production of these diverse and complex secondary metabolites (Bibb *et al.*, 1989, see also sections 1.4.3 and 1.4.4).

Orf11 and Orf13 are homologous to RedY and RedV respectively, which encode proteins believed to be involved in undecylprodigiosin biosynthesis

in *Str. coelicolor* A3(2). Since both proteins in *S. marcescens* and *Str. coelicolor* A3(2) do not have homology to any other database proteins, it is likely therefore, that they are unique to prodigionine biosynthetic pathways.

Orf12 is weakly homologous to the *psf* gene product of *Bacillus pumilis*. In *B. pumilis*, this protein is required for surfactin antibiotic production where it has been suggested to be a possible regulator (Morikawa *et al.*, 1992). Other members of the family to which this protein belongs include Sfp, involved in surfactin biosynthesis in *B. pumilis* (Nakano *et al.*, 1992), and enterobactin synthetase (EntD) which is involved in iron chelation in *E. coli* (Coderre *et al.*, 1989). Further analysis of this homologue in comparison to Orf12 should reveal where the similarities are in relation to possible function, and this is particularly important in this instance as homology is not strong. However, the possibility of a putative regulatory protein being present in the *pig* cluster opens up many exciting avenues of potential investigation in the future.

Orf15 and Orf16 were found to have database homologues which were identified during various genome sequencing projects (see Table 4.5) and are therefore uncharacterised at present. Consequently, the possible roles of these proteins in prodigiosin biosynthesis remains unknown.

Based on the findings presented in this section, and on information obtained from database accession sources, a comparison of pigment gene ordering between *S. marcescens* and *Str. coelicolor* A3(2) showed that the organisation of the *pig* cluster differs from the *red* cluster. This raises interesting issues concerning the possible source of these genes and how the two species could have acquired biological pathways to produce metabolites of no immediately obvious function, which are similar in many aspects yet considerably different in others; for example, is this suggestive of convergent evolution or of horizontal transfer? One fact which must be considered is that since Red is a mixture of prodiginines (discussed in section 1.3.5.1), and the *red* biosynthetic pathway has already been predicted to be more complex in *Str. coelicolor* A3(2) than that of prodigiosin biosynthesis in *S. marcescens*, the differences in gene order in these clusters is also likely to be a reflection of this.

4.10 GENERAL DISCUSSION

Many of the putative proteins deduced from translation of the nucleotide sequence of the *pig* cluster were found to be homologous to database proteins. This information provided further evidence to support the analyses which had been done to assign putative ORFs to the sequenced region, i.e it demonstrated that most putatively assigned coding regions do indeed encode proteins in theory. However, it was found - particularly regarding the results of multiple sequence alignments - that there are also many differences between the putative prodigiosin biosynthetic proteins and their respective database homologues. It is noted that putative proteins identified in this study often had, possible undetected sequencing errors notwithstanding, longer amino acid sequences than their database counterparts.

Computer analysis of putative proteins reinforces the notion that although extremely useful as a first step, it is not possible to assign ORFs, and in particular their translation initiation sites, to sequenced regions of DNA with a high degree of certainty on the basis of the sequencing data alone. Translation initiation sites and sequences of encoded proteins, need to be confirmed physically. At the protein level, N-terminal sequencing would provide unequivocal evidence of translational start sites. Due to time constraints, these experiments were beyond the scope of this study.

In order to confirm that all of the products of the putative ORFs assigned to the sequenced region are required for prodigiosin biosynthesis, a directed mutagenesis strategy should ideally be employed. Comprehensive sequence data, which were unavailable at the outset of this study, are essential for this, because it would be important to ensure that any insertion or deletion mutation introduced maintains the reading frame because of the possible polar effects on downstream genes. In this study, random transposon mutagenesis was successfully used to mutate *orf1* and *orf4*,(described in subsequent chapters) but

with the sequence data derived from this and the parallel study, a more directed approach would be the obvious one.

For those proteins which have been tentatively predicted to be targeted to the cell membrane (in this study, Orf1, Orf2 and Orf3), there is much scope for further work to confirm or disprove these predictions: more analysis of the likelihood of each of these putative proteins possessing transmembrane domains is required, to build on preliminary data generated during this study. One way in which this could be achieved is by more detailed computer analysis of secondary structure and hydropathy, but ultimately, toplogy probe analysis of each hypothetically cell membrane targeted protein would be required.

Genetic approaches to determine the topology of cell-membrane proteins have been successfully used in other systems, for example in characterising the Out protein components of the *Erw. carotovora* subsp.*carotovora* general secretory apparatus (Reeves *et al.*, 1993), the haemolysin secretory protein HlyB in *E. coli* (Wang *et al.*, 1991) and penicillin-binding-protein 1B in *E. coli* (Edelman *et al.*, 1987). The toplogy probe is usually reporter gene-encoded mature β-lactamase or alkaline phosphatase (Broome-Smith *et al.*, 1990; Manoil *et al.*, 1990; Manoil and Beckwith, 1985). In this study a random transposon-based strategy (described in Chapter 5), was attempted to try to deduce the cellular locations of *pig* cluster-encoded proteins.

In summary, work presented so far in this chapter, together with the work done by A. Cox, forms the basis for many further avenues of investigation. As well as further studies on the molecular genetics of prodigiosin production, future work should also focus on the cellular localisation and biochemistry of pigment biosynthesis. To begin the physical characterisation of the putative protein products identified in this study, *in vivo* gene expression experiments were carried out and this work is the subject of the next section of this chapter.

4.11 *IN VIVO* EXPRESSION OF THE PUTATIVE PROTEINS ENCODED BY THE 5' END OF THE *pig* CLUSTER IN THE T7 SYSTEM

4.11.1 PREFACE

In order to assess if the putative proteins from the 5' end of the *pig* cluster are functional *in vivo*, their DNA coding regions were subcloned and expressed in the T7 system (Tabor and Richardson, 1985). This expression system was chosen because it does not require the gene of interest to have a functional promoter, which is important in this context, because ORFs are predicted to be translationally coupled. Although computer analyses had shown that the coding regions are likely to be genes and database homologues had been identified for the majority of the putative proteins, these data remain hypothetical unless a protein product can be visualised in an expression system. This is particularly important with regard to Orf4, which does not have any database homologues and is therefore likely to be a novel protein.

In the T7 system, the gene of interest is cloned into vector pT7-5 or pT7-6 (Appendix I), such that its expression is under the control of the T7 polymerase promoter. These plasmids only differ from one another by the orientation of the multiple cloning site with respect to the T7 promoter. The activity of the promoter is under the control of T7 RNA polymerase which is encoded by gene *I* of the phage T7, carried separately on a second vector, pGP1-2. pGP1-2 is constructed such that the RNA polymerase gene expression is dependent upon activity of the inducible λP_L promoter and the λc_{I857} temperature-sensitive repressor. Following induction by temperature-shift to 42°C, exclusive expression of the gene of interest should be achieved by the addition of rifampicin (Rf), an antibiotic which specifically inhibits host-encoded RNA polymerase by the formation of a very stable complex with the molecule (Wehrli *et al.*, 1968a, 1968b). Transcription of host genes should therefore be prevented and since T7 RNA polymerase is unaffected by Rf, the gene of interest is still expressed under the direction of the T7 promoter. Protein products are labelled with ^{35}S-Methionine, separated by SDS-PAGE and visualised by autoradiography as described in section 2.7.4.

4.11.2.1 CLONING OF THE 5' END OF THE *pig* CLUSTER INTO pT7-6

pT7-6 and the host strain *E. coli* K38(pGP1-2) were provided by N. Thomson and G. Bosgelmez respectively. Subcloning of the 5' end of the *pig* cluster was made difficult by the limited choice of restriction sites in the multiple cloning sites of pT7-5 and pT7-6 and also, in relation to this, the lack of corresponding, conveniently located restriction sites in the *pig* cluster due its size.

In an ideal T7 expression study of ORFs of this number and apparent organisation (clustered), one would endeavour to clone several of the ORFs together, in their respective transcriptional units. Downstream translationally coupled genes may not be expressed if cloned singularly, due to inefficient RBSs, if separated from the first gene of the transcriptional unit. Also, constructs should be devised whereby expression would produce truncates of some of the encoded proteins, the sizes of which would be predictable by the restriction enzyme cut site in the DNA sequence, therefore providing further physical evidence to confirm predictions about the coding regions.

Attempts to construct progressive ordered truncates of each ORF were hampered by having to use blunt-ended cloning techniques and having to recover DNA between sequential digestions due to incompatibility of restriction enzyme buffer requirements. Poor DNA recovery due to successive treatments inevitably resulted in reduced efficiency, compounded by the inherent inefficiency of blunt-ended ligation strategies. As a consequence, only two recombinant plasmid constructs (pT7-6PB and pT7-6PE) made during this study and they are represented in Fig. 4.13. A further consideration for cloning experiments and choice of enzymes for digestion was that since fragments were to be isolated from pM245, fragment size predictions for a digestion experiment had to be unequivocal and it had to be ensured that fragments generated from digestion of plasmid pM245 were of sufficiently differing sizes to allow resolution on an agarose gel so that the correct fragment is ultimately cloned in each case.

4.11.2.2 GENERATION OF PLASMIDS pT7-6PB and pT7-6PE

Construction of both pT7-6PB and pT7-6PE required the use of pT7-6 because the polylinker is in the correct orientation with respect to the promoter in each case. pT7-6PB was constructed by unidirectional cloning of the 5.415Kb *Pst*I-*Bam*HI fragment from pM245 into pT7-6; pT7-6PE was constructed by unidirectional cloning of the 12.766Kb *Pst*I-*Eco*RI fragment from pM245 into pT7-6. A representation of the cloning strategy is given in Fig. 4.13.

The *Pst*I-*Bam*HI and *Pst*I-*Eco*RI insert fragments were obtained from pM245 by double digestion (section 2.3.4.1) and isolated from an agarose gel after electrophoresis by Geneclean™ (BIO101, section 2.3.4.7). DNA fragments were ligated to *Pst*I-*Bam*HI digested, or *Pst*I- *Eco*RI digested, alkaline phosphatase treated, pT7-6 (sections 2.3.4.1, 2.3.4.2 and 2.3.4.4) and introduced into *E. coli* DH5α by $CaCl_2$ transformation (section 2.4.1). Plasmid recombinants were selected on NBA-Ap. Recombinant constructs were extracted by mini-prep alkaline lysis (Qiagen, section 2.3.1) and verified by digestion with *Pst*I/*Bam*HI or *Pst*I/*Eco*RI followed by agarose gel electrophoresis against 1Kb Ladder size standards.

Recombinant plasmids pT7-6PB and pT7-6PE were introduced into *E. coli* K38(pGP1-2) by electroporation (section 2.4.2) and transformants selected on NBA-Ap-Kn. It was known from previous work that the gene products encoded by *S. marcescens* DNA carried on pM245 were not toxic to *E. coli* (A. Cox, Pers. comm.), despite the high copy number of this plasmid. Approximately half of the *pig* cluster is present in the *Pst*I/*Bam*HI fragment and this was cloned with ease into pT7-6, therefore it seemed favourable that this construct would not have any deleterious effect in the *E. coli* K38(pGP1-2) host strain. This was an important observation to make prior to the expression experiment as many genes produce mild toxic effects when cloned into K38(pGP1-2) (Tabor, Pers. comm.) In order to minimise any potential harmful effects in this host strain, the transformation procedure was carried out no more than a few days before the induction experiment was done. In addition to this, a control electroporation experiment was done whereby *E. coli* K38(pGP1-2) was transformed with pT76 vector. No significant difference was seen in the number of transformants obtained in

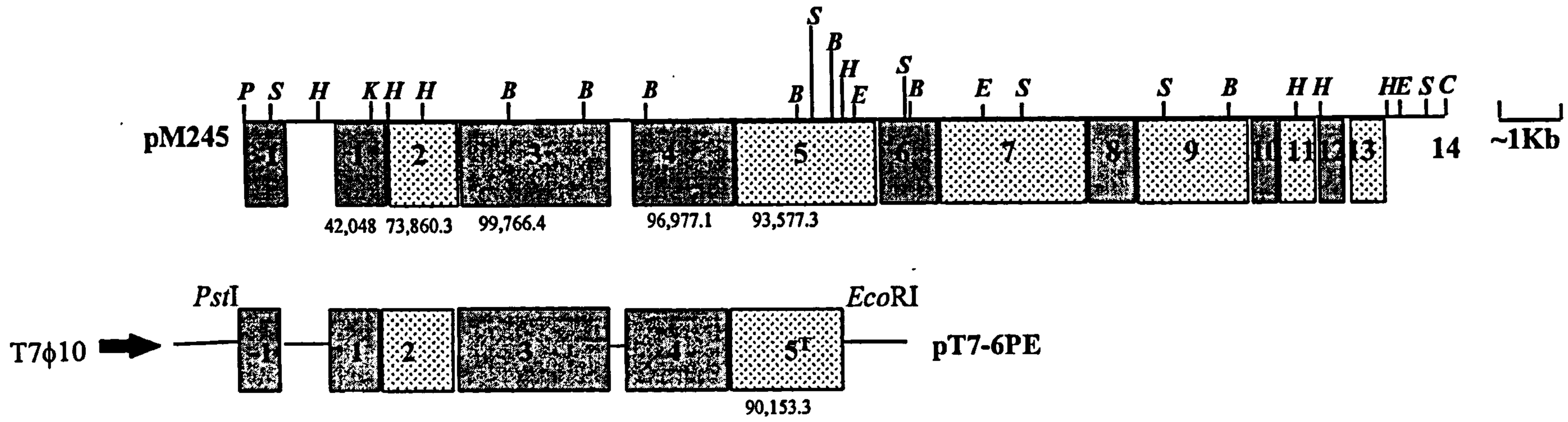


FIGURE 4.13 pT7-6 DERIVATIVES CONSTRUCTED FOR *IN VIVO* EXPRESSION STUDIES

ORFs are represented by shaded boxes. Numbers refer to ORF designations. Predicted molecular weights (Da) of the encoded proteins are shown below boxes where relevant. These predictions are based on the assumption that there is no read-through the restriction site until a transcriptional or translational terminator is reached in the fusion proteins constructed in these recombinant plasmids. Such a possibility cannot be discounted for certain given the presence of the ß-lactamase casette in the vector

See text in section 4.11 for details.

Enzymes Key: **B**:*Bam*HI; **C**:*Sac*I; **E**:*Eco*RI; **H**:*Hin*dIII; **K**:*Kpn*I; **P**:*Pst*I; **S**:*Sal*I

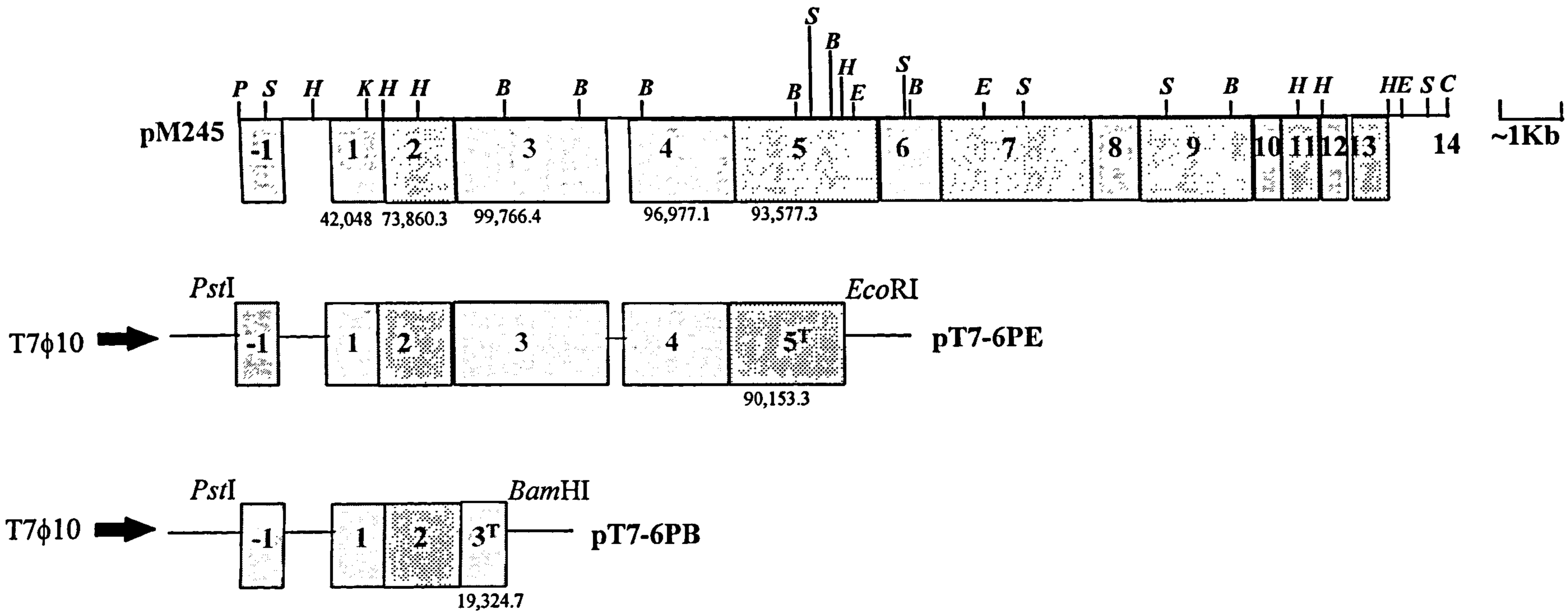


FIGURE 4.13 pT7-6 DERIVATIVES CONSTRUCTED FOR *IN VIVO* EXPRESSION STUDIES
ORFs are represented by shaded boxes. Numbers refer to ORF designations. Predicted molecular weights (Da) of the encoded proteins are shown below boxes where relevant. See text in section 4.11 for details.

Enzymes Key: **B**:*Bam*HI; **C**:*Sac*I; **E**:*Eco*RI; **H**:*Hin*dIII; **K**:*Kpn*I; **P**:*Pst*I; **S**:*Sal*I

comparison to transformations with pT76-PB and pT76-PE. A marked reduction in transformation efficiency of this strain with either of the recombinant plasmids would have suggested that the encoded *S. marcescens* gene products are toxic to the host and would therefore not be produced in this strain.

4.11.3.1 *IN VIVO* EXPRESSION OF pT76-BASED CONSTRUCTS

The expression experiment was performed as previously described (section 2.13), and the resulting radiolabelled proteins separated by SDS-PAGE (section 2.14.1). Figure 4.14 shows the results of the induction experiment. A standard curve (data not shown) was obtained by plotting log distance migrated (mm) against size (KDa) of the markers used for SDS PAGE (section 2.14.2) and was used to estimate the sizes of the band obtained following SDS PAGE and auto-radiography.

4.11.3.2 RESULTS AND DISCUSSION OF THE pT76-PB EXPERIMENTS

A comparison of lanes 1B and 1C with lane 1A in Fig. 4.14 shows that rifampicin effectively prevented the expression of host-encoded proteins following the temperature shift. In lane 1A, it can be seen that there is one highly visible band and two fainter bands. Using the standard curve, the sizes (KDa) of these bands were estimated to be 48.9, 33.1 and 30.2. The predicted sizes (KDa) of the proteins which should be present in this construct (shown in Figure 4.13) were: Orf1: 42.042; Orf2 (unprocessed): 73.860, and Orf3 (truncated): 19.325.

Results of the T7 experiments do not therefore correspond to the sizes predicted for the ORFs in this construct. It is possible that the 48.98KDa band in lane 1A is Orf1 and due to possible aberrant mobility, it appears to be larger than it actually is. Another possibility is that the predicted transcriptional start site of *orf1* is not in fact the true start site. The possibility of alternative start sites for this ORF was discussed in section 3.4.1; however, even if the start site were several residues upstream of the designated methionine codon (see Fig, 3.5b), this would only make a marginal difference to the estimated size of the protein, which in this case would be predicted to be 42.89KDa. This is still considerably smaller than

FIGURE 4.14 AUTORADIOGRAPH SHOWING THE RESULTS OF T7 EXPRESSION EXPERIMENTS

The lanes on the gel are in two sets of three. Those prefixed (**1**) show results of expression experiment using the pT76-PB construct; those prefixed (**2**) show results of expression experiments using the pT76-PE construct.

From the left hand side, the lanes represent the following:
M: Rainbow™ ^{14}C Methylated protein molecular weight markers (see section 2.14.)
1A:pT7-6PB, 30°→42°C induced +Rf
1B:pT7-6PB, 30°→42°C induced -Rf
1C:pT7-6PB, 30° uninduced
2A:pT7-6PE, 30°→42°C induced +Rf
2B:pT7-6PE, 30°→42°C induced -Rf
2C:pT7-6PE, 30° uninduced

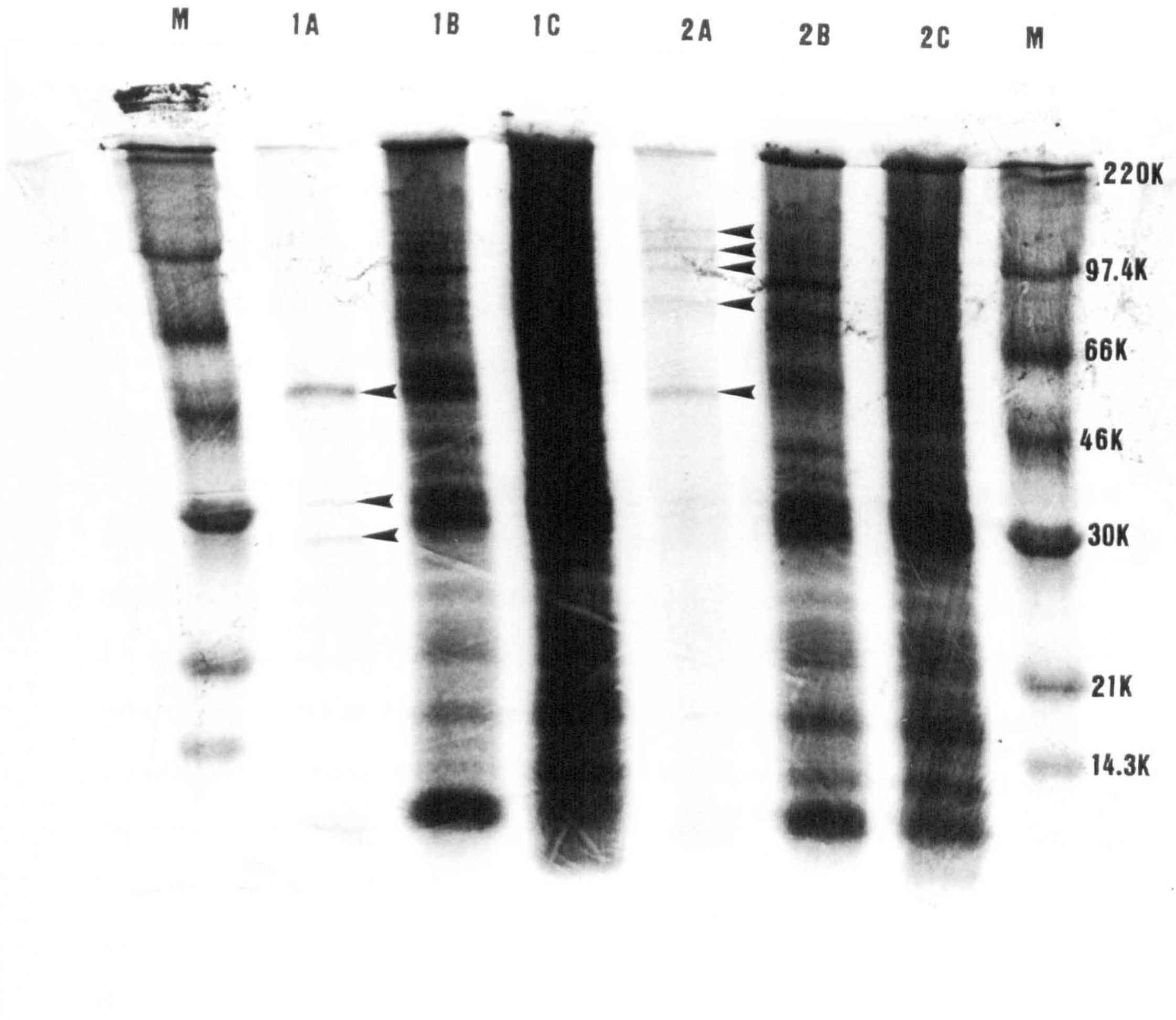
Molecular weights of the markers are given next to lane **M** on the right-hand side of the autoradiograph.

Lanes marked **C** result from uninduced cells, and show cellular background proteins from genes expressed by host-encoded RNA polymerase.

Lanes marked **B** result from induced cells following the temperature shift to 42°C in the absence of rifampicin and show those cellular background proteins that are highly expressed by host-encoded RNA polymerase together with the proteins encoded by the cloned genes and expressed from the T7 promoter by the temperature-induced T7 RNA polymerase.

Bands marked by arrows are discussed in the text, see sections 4.11.2.2 and 4.11.2.3 for interpretation of results.

Figure 4.14



the estimated band size according to the autoradiograph. However, the fact that this band is highly visible in lane 1A (and also lane 2A) is interesting because it is predicted that Orf1 may be well expressed *in vivo*, firstly because it is the first gene in the putative transcriptional unit, and secondly because it hypothetically posesses a good putative ribosome binding site. Also, evidence from gene fusion studies (presented in Chapter 7) supports the hypothesis that Orf1 is relatively well expressed in the cognate host system.

The two fainter bands in lane 1A, that are estimated to be ~33 and ~30 KDa, do not correspond to estimated sizes of Orf2 and Orf3. As stated earlier (section 4.5.1), Orf2 is predicted to have a signal peptide cleavage site; the estimated size of the protein if it were processed is 71.76KDa, and since this is only marginally smaller than the predicted size of the unprocessed form, it does not seem likely that either of the two smaller bands visible in lane 1A could represent Orf2. Neither of the two smaller bands can represent a truncated Orf3, as it seems extremely improbable that possible aberrant mobility of this protein would result in such gross retardation during electrophoresis.

Another possible explanation for the two smaller bands in lane 1A is that they are artefactual degradation products which arise due to these genes being expressed in the *E. coli* background. Assuming that the two smaller bands in lane 1A are artefactual does not account for the apparent absence of bands for Orf2 and the truncated Orf3 at the predicted sizes. One possible explanation is that translational coupling of these genes is having a detrimental effect on their expression, i.e. they are being expressed at very low levels and consequently fail to show. Artefactual bands can also theoretically arise when a subpopulation of protein molecules migrate abberrantly when they are not fully denatured. However, the same results were obtained for the T7 experiments when the experiment was repeated (data not shown), so the results are reproducible. Results of T7 experiments might be suggestive of the possibility that predictions for encoded proteins are not correct: as discussed in previous sections, there remains the possibility that Orf2 (and Orf5) might represent more than two proteins but unfortunately, this cannot be ascertained without further work which was beyond the scope of this project..

4.11.3.3 RESULTS AND DISCUSSION OF THE pT76-PE EXPERIMENTS

As with the pT76-PB experiments, comparison of lane 2A with lanes 2B and 2B in Fig. 4.14 shows that expression of host-encoded proteins was effectively prevented in this experiment by the addition of rifampicin. The bands which have been marked in lane 2A were estimated, from the standard curve, to be of the following sizes (KDa): 116.1, 97.7, 91.2, 74.1 and 49. The predicted sizes (KDa) of the proteins in this construct are shown in Fig. 4.12 and are as follows: Orf1: 42.048, Orf2: 73.89, Orf3: 99.766, Orf4: 96.977, and Orf5 (truncated): 90.153.

In lane 2A, it can be seen that three very high molecular weight bands have migrated close to one another. The size of the highest band in the lane is greater than that which is predicted for Orf3, which is the largest putative protein in the construct. The size estimate of the second largest band in the lane corresponds quite well to the predicted size of Orf4, differing by 743Da. The third largest band in the lane could correspond to the slightly truncated Orf5 protein present in this construct. It is possible that these bands appear to be larger than they actually are because of several possible reasons. As these bands are co-migrating due to size similarity, and this might be affecting their mobility. One or more of these bands might have aberrant mobility through the gel, or they might have been slightly retarded during electrophoresis because of the number of proteins in the track. The band estimated to be 74.13KDa in lane 2A could represent Orf2, being only slightly larger than the predicted size for this protein. The size of this band is closer to the estimated size for the unprocessed protein, but it is difficult to tell by looking at the gel if a co-migrating band is present which would represent the processed protein. It is interesting that this protein was visible in expression of this construct, which is much larger than pT76-PB and this raises the possibility that it is an artefact in this experiment. Once again, a band of an estimated 48.98KDa is seen also in this lane and this may represent Orf1 migrating at a slower rate than would be predicted (discussed in section 4.11.3.2). It can also be seen that this band is also the strongest one in this experiment, so if it does represent Orf1, this supports the prediction it would theoretically be most srongly expressed.

4.11.4 SUMMARY OF T7 EXPRESSION STUDIES

The work described in Section 4.11 was partly successful in tentatively showing that when *S. marcescens* DNA from the 5' end of the *pig* biosynthetic cluster is put into an expression system, protein products can be visualised. The sizes of the bands resulting from expression of the putative genes do not precisely match the predictions of the molecular weights for these proteins, but they are, for the most part, broadly within the predicted size range for each protein. Since other results obtained (presented in earlier sections of this chapter), are, with one exception, strongly suggestive of these ORFs encoding proteins with known homologues, the results obtained in this section are more acceptable as preliminary experimental data.

Clearly, further expression studies of this kind are required in order to produce unequivocal physical evidence of the nature of the proteins encoded in the 5' end of the *pig* cluster. Future strategies, so far as T7 expression work is concerned, would require circumvention of the strategical problems encountered in subcloning such a large fragment of DNA, i.e. the lack of useful yet exclusive restriction enzyme sites in *pig* cluster DNA. The aim would be to produce subclones of each transcriptional unit which are progressively smaller so that a full size and truncated protein is expressed for each ORF. Due to the likely translational coupling, it is predicted that *orf2*, *orf3* and *orf5* may be poorly expressed in T7 experiments, if at all, if they were cloned singularly. The problem of finding sufficiently rare yet compatible restriction sites for cloning into the limited multiple cloning site of the T7 vectors for a given subclone could be overcome in the future by using oligonucleotide primers designed to specific regions of *pig* cluster DNA which incorporate desired enzyme sites and performing PCR.

4.12 GENERAL SUMMARY

Work presented in this chapter was that which was done to build upon the raw nucleotide sequence data which was presented in Chapter 3. It provides information on the likely characteristics of the gene products encoded in the 5' end of the *pig* cluster in *S. marcescens*. Database searches revealed that many of the

putative genes identified in this study and parallel work have much in common with those located in and around the undecylprodigiosin biosynthetic cluster of *Str coelicolor* A3(2), lending support to biochemical and genetic cross-feeding studies done between these species previously (see section 1.3.5.1). *In vivo* expression studies showed that subcloned regions do encode protein products. The work presented in this chapter provides a basis for many future experiments which were, due to time constraints, beyond the scope of this study, including studies on the biochemistry of encoded proteins and their cellular localisation. For those putative proteins for which database homologues do not exist at present, it will be interesting to see if homologues are found in the future, as more bacterial genomes are systematically sequenced and their gene products characterised.

As work presented in Chapters 3 and 4 was underway, the genetics of pigment production was studied concurrently from another angle. In the absence of comprehensive sequence data, random transposon mutagenesis strategies were used to gain further insight into the genetics of prodigiosin biosynthesis. This work is described in Chapter 5.

CHAPTER 5

TRANSPOSON MUTAGENESIS OF THE PRODIGIOSIN BIOSYNTHETIC CLUSTER

5.1 PREFACE

As work was progressing to sequence the *pig* cluster, experiments were also performed to isolate pigment biosynthetic mutants. This work is presented in this chapter. A random mutagenesis approach was used because there was insufficient sequence information to use a more directed method of mutagenesis.

5.2 AIMS

The main aim of the work described in this chapter was to generate *S. marcescens* mutants which were altered or abolished in prodigiosin biosynthesis by using a random "transposon tagging" approach, which would allow isolation of genes involved in prodigiosin biosynthesis, and also generate gene fusion strains. Another aim of one particular strategy, the use of the transposon Tn*blaM*, was to see if gene fusions using this transposon could be obtained, as this would provide physical evidence of cellular locations of the biosynthetic proteins.

5.3 THE USE OF PHAGE λ-BASED TRANSPOSON TECHNOLOGIES IN *Serratia marcescens*

A common approach in the genetic dissection of biosynthetic pathways is the generation of mutants affected in synthesis of the product under investigation. There are many possible approaches which can be used to generate mutations and the molecular analysis of these mutants allows specific genetic loci to be assigned to particular functions in biological systems. Methods of inducing random mutations in bacterial genomes include chemical (EMS and NTG), physical (UV-irradiation), and genetic (transposon mutagenesis) methods.

Previous work done in the laboratory had demonstrated that *S. marcescens* ATCC39006 and its derivatives could be made susceptible to coliphage λ by introduction of a *lamB* carrying plasmid. The use of λ as a delivery vector for transposons in enteric strains other than *E. coli* has been used with success in studying the molecular genetics of exoenzyme production in the closely related species *Erw. carotovora* subsp. *carotovora* (Salmond *et al.*, 1986). For example, *Erw. carotovora* subsp. *carotovora* strain SCRI193 was successfully rendered susceptible to λ infection by transformation with plasmid pHCP2

(Clement *et al.*, 1982), carrying a cloned version of the *E. coli lamB* gene. This gene encodes an outer-membrane bound protein which serves as the λ receptor (Randall-Hazellbauer and Schwartz, 1973; Schwartz and Le Minor, 1975). The presence of *lamB* in a multicopy plasmid vector allows phage adsorption and DNA injection, but lytic growth is suppressed due to blocked replication, morphogenesis, or transcription (Schwartz and le Minor, 1975, de Vries *et al.*, 1984, Harkki and Palva, 1985). *lamB* therefore allows λ to be used as a suicide vector for the transposon carried in the construct.

Mutagenesis of *S. marcescens* in order to "tag" pigment biosynthetic genes was greatly facilitated by the fact that the defective phenotype is easily screened for, by visual comparison with wild-type (pigmented) colonies following mutagenesis. In addition to the use of λ as a suicide vector, the strategy used in these experiments has other advantages. At the molecular level, the advantages of using transposon-based technologies are manifold, because transposons introduce a large amount of extraneous information into the host system which can be exploited. Firstly, they carry an antibiotic resistance marker so when a mutant with the desired phenotype is obtained, the gene of interest can be located. The mutation can be visualised by Southern blot hybridisation, providing a direct route to cloning the flanking DNA, ultimately enabling identification of the disrupted gene. The mutation can be moved into different genetic backgrounds by using, where it exists, generalised transducing phage technology and exploiting the co-inheritance of the selectable marker with the inactivated gene. Transposons which have been constructed to carry reporter genes allow, by a single mutagenesis experiment, the creation of gene fusion strains which no longer possess the wild-type characteristic and produce instead an enzyme whose activity is readily assayable by simple biochemical methods in the laboratory. Such strains can therefore be used to study gene expression in response to varying environmental stimuli.

The two transposons which were used in the isolation of non-pigmented mutants in this study shall be described in turn, and they are depicted in Fig. 5.1. The direct strategy for obtaining reporter gene fusions in relevant genes involved the use of λTn*phoA*'-2. This composite transposon is a derivative of

Figure 5.1

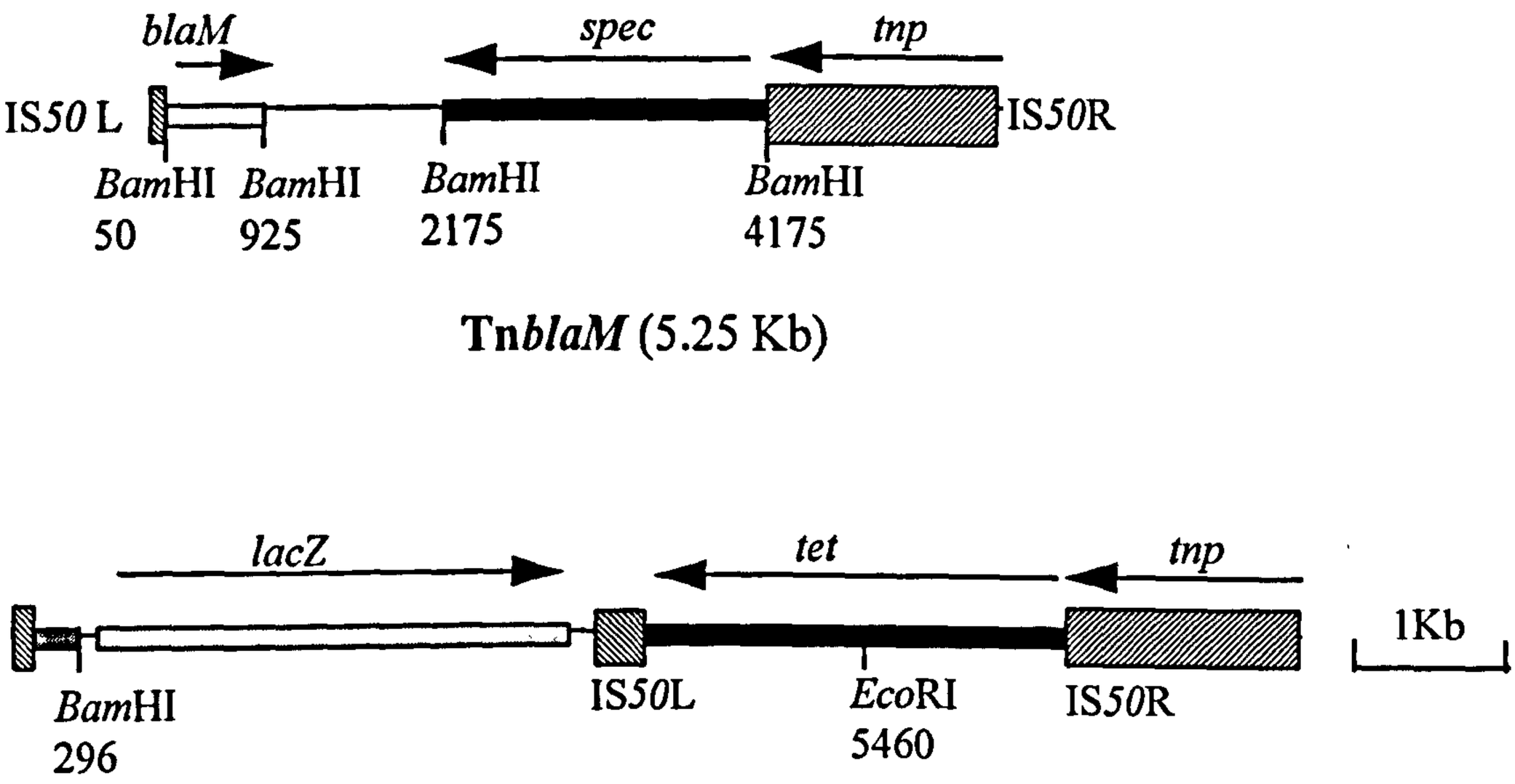


FIGURE 5.1 THE STRUCTURES OF COMPOSITE TRANSPOSONS Tn*blaM* AND Tn*phoA*'-2

Structural organisation of the transposons, and relevant restriction sites are shown in each case, with their positions in nucleotides in the DNA sequence of the transposon. After Taddyon and Broome-Smith, 1992 and Wilmes-Reisenberg and Wanner, 1992.

Tn*phoA*, where the *phoA* gene, encoding mature bacterial alkaline phosphatase has been replaced with the *lacZ* gene encoding β-galactosidase and the selectable marker is Tc (Wilmes-Reisenberg and Wanner, 1992). The *lacZ* gene is promoterless, but has a ribosome binding site, so it was used in this study to create transcriptional fusions. The insertion sequences (IS) elements in Tn*phoA*'-2 are IS*50*L and IS*50*R, and they are essentially the same as those of Tn*5* (Berg and Berg, 1983). *lacZ* is actually fused to the N-terminal encoding region of *phoA* from the original parental construct Tn*phoA*. The point of fusion is base 240, corresponding to amino acid residue 73 in mature PhoA. The greater portion of IS*50*L is located downstream (3') of the reporter gene cassette in Tn*phoA*'-2, and Tc^R is located between IS*50*L and IS*50*R (Fig. 5.1)

The other transposon-based strategy which was employed was mutagenesis using λTn*blaM* (Taddyon and Broome-Smith, 1992, Fig. 5.1). This transposon was constructed to provide an simple but elegant way to tag bacterial genes encoding cell envelope and secreted proteins. Tn*blaM* is derived from Tn*5*. In Tn*5*, transposable elements IS*50*L and IS*50*R lie either side of a central region encoding Kn^R (Berg and Berg, 1983). Only IS*50*R encodes transposase as IS*50*L contains a stop codon within its coding region. In addition to the outer terminus and transposase-coding region of IS*50*R, only the terminal 19bp of IS*50*L are required for efficient transposition of the intervening sequences of Tn*5*, so in Tn*blaM*, all but the essential 19bp of IS*50*L are replaced by the mature β-lactamase (BlaM) coding region. The Kn^R marker of Tn*5* is replaced by Sp^R in this construct (Taddyon and Broome-Smith, 1992, Fig. 5.1); *blaM* carried in the transposon is devoid of a ribosome binding site and expression of this gene is therefore dependent upon the transposon being inserted in the correct orientation *and* in- frame with respect to the transcriptional direction of the disrupted host gene. In order to obtain a phenotypically $BlaM^+$ (Ap^R) fusion, Tn*blaM* must be inserted into a secreted protein encoding gene, or an integral membrane protein encoding gene. In the latter case, the point of fusion must produce a chimaeric protein where the mature β-lactamase component is present in a periplasmic domain.

The rationale behind using Tn*blaM* to mutagenise *S. marcescens* was to obtain prodigiosin biosynthetic mutants and assess them for β-lactamase activity (Ap^R). Any Ap^R colonies obtained which are also altered for pigment production, would provide evidence that the affected prodigiosin biosynthetic genes encode periplasmic or integral-membrane proteins.

5.3.2 ISOLATION OF PIGMENT MUTANTS USING λTn*phoA*'-2

Previous problems which had been experienced with genetic manipulation of the progenitor ATCC39006 strain, apparently due to its restriction-modification system (N. Thomson, Pers. comm.; A. Cox. Pers. comm.), were overcome by the use of two derivative strains - SCAR1310124 and SCAR952, constructed by A. Cox. SCAR1310124 was made by λTn*10* mutagenesis followed by fusaric acid "curing" (Bochner *et al.*, 1980) in order to excise Tn*10* from the chromosome. SCAR952 was used in λTn*blaM* mutagenesis experiments and shall be discussed in section 5.3.2. SCAR1310124 was made susceptible to λ infection by transformation with plasmid pMUT13, which carries a Kn^R marker (Appendix I).

The problem of screening SCAR1310124-derived transductants for potential gene fusions (as this strain is Lac^+) was circumvented by the use of the *S. marcescens* generalised transducing phage ΦOT8 which was isolated and characterised in this laboratory (S. Bentley, unpublished; W. Orme and N. Thomson, unpublished; Cox, 1995), to transduce the mutations at a later stage in the study into a Lac^- strain (SLAC1) which was generated as part of this study (described later in this chapter). Generalised transduction experiments are described in Chapter 6.

Tn*phoA*'-2 was introduced into SCAR1310124 (pMUT13) by λ transduction using a high-titre lysate, (10^9-10^{11} $pfuml^{-1}$, section 2.5.1, original lysate provided by M. Sebaihia), as described previously (section 2.5.2). The experiment was carried out several times on overnight cultures of the strain, each of which had been grown from a single colony. This ensured that any potentially interesting transductants were independent isolates and not siblings. Non-transduced aliquots of SCAR1310124 (pMUT13) were also plated on NBA-Tc

and incubated as a negative control for each experiment. Following transduction and expression, transductants were selected on NBA-Tc. Plates were incubated at 30 C for three days and any mutants altered in their pigment phenotype, by visual comparison with wild-type, were retained for further study

Mutants visually differing from wild-type for pigmentation were replicated onto NBA without selection as a master plate because maintaining selection for the transposon-encoded resistance marker increases selection pressure for secondary transposition events. Colonies were also replicated onto the following plates: MMA to ensure prototrophy, and carbapenem assay plates seeded with *E. coli* ESS (section 2.9) to ensure that they are not *rap* mutants (see section 1.3.4.2). All these plates were incubated at 30°C for three days, with the exception of carbapenem bioassay plates which were incubated at 25°C overnight In addition to obvious pigment mutants, at least 200 other transductant colonies per experiment were also replicated onto NBA and MMA plates to ascertain a frequency rate for auxotrophy as an indication of the efficiency of the mutagenic process.

5.3.2 ISOLATION OF PIGMENT MUTANTS USING λTn*blaM*

SCAR952 (pTROY9) was constructed by A. Cox in order to overcome the difficulties which had been previously experienced with genetic manipulations of the progenitor strain ATCC39006. SCAR952 was created by Tn*5* mutagenesis, and retains this transposon in its chromosome because it cannot be cured using fusaric acid. Use of λTn*blaM* in this strain overcame other technical problems: a LamB-encoding plasmid other than pMUT13 had to be used because this encodes Kn^R so this strain was transformed with pTROY9 (deVries *et al.*, 1984). pTROY9 encodes Tc^R, therefore this strain SCAR952 is not suitable for λTn*phoA*'-2 mutagenesis.

λTn*blaM* was provided by M. Sebaihia. A high titre lysate of this (>10^9 pfu ml^{-1})was made on the suppressing strain *E. coli* LE392 (Table 2.1) as described previously (section 2.5.1) and Tn*blaM* was introduced into SCAR952(pTROY9) by λ transduction (section 2.5.2). The experiment was carried out on several cultures, each of which had been grown from single

colonies. Following transduction and expression, cultures were plated out on NBA-Sp plates and incubated for three days at 30°C. As negative control, a non-transduced aliquot of the culture was also plated on NBA-Sp for each experiment. Colonies with pigment phenotypes differing from wild-type were retained for further study.

Mutants visually differing from wild-type were replicated onto NBA master plates without selection. These colonies were also replicated onto the following plates: MMA to ensure prototrophy, carbapenem bioassay plates seeded with *E. coli* strain ESS to ensure that they are not *rap* mutants (see section 1.3.4.2) and NBA-Ap plates to screen for BlaM fusions. All plates were incubated at 30°C for three days, with the exception of carbapenem bioassay plates which were incubated at 25°C overnight.

5.4.2 RESULTS OF MUTAGENESIS EXPERIMENTS

Results of the mutagenesis experiments for each strain are presented in Table 5.1. The appearance of different types of pigment mutants on NBA, obtained from a typical transposon mutagenesis experiment is shown in Fig. 5.2. Each pigment-mutant strain which was analysed further was confirmed to be prototrophic by growth on MMA. Strains which were generated from these experiments are listed in Table 2.1.

TABLE 5.1 RESULTS OBTAINED FROM λTn*phoA*'-2 MUTAGENESIS OF SCAR1310124 and λTn*blaM* MUTAGENESIS OF SCAR952

	% MEAN FREQUENCY PER TRANSDUCTION[a]:	
MUTANT PHENOTYPE	**SCAR952(pTROY9) ::Tn*blaM***	**SCAR1310124 (pMUT13)::Tn*phoA*'-2**
Non-pigmented (Pig^-)	0.5	0.3
Hyperpigmented (Pig^{++})	0.2	0.1
Orange (Pig^O)	1	7.8
Auxotrophic	1	1.4

a: Frequencies were derived by comparing the number of each mutant phenotype with the total number of colonies screened. Mean total numbers of colonies per transduction were: Tn*blaM*, 202; Tn*phoA*'-2, 768

Figure 5.2

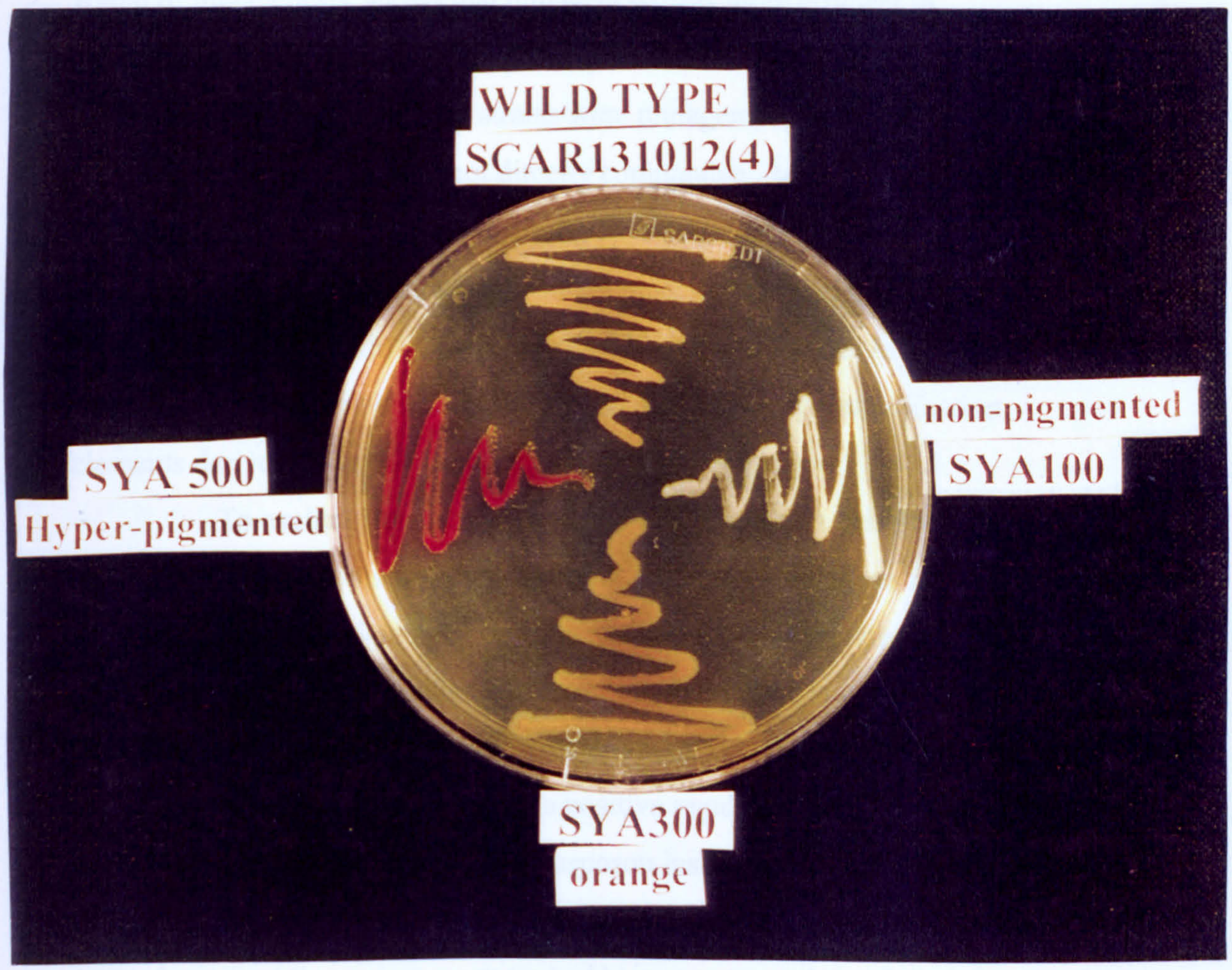


FIGURE 5.2 PIGMENT PHENOTYPES OBTAINED BY TRANSPOSON MUTAGENESIS OF *Serratia marcescens*

The transductants shown above, with a wild-type control were grown on NBA. These transductants resulted from λTn*phoA*'-2 mutagenesis of *Serratia marcescens* SCAR1310124 (pMUT13), but the phenotypes are typical of every transposon mutagenesis experiment carried out in this study.

5.4.3 DISCUSSION

Auxotrophy frequencies of 1% (Tn*blaM*) and 1.4% (Tn*phoA*'-2) indicated that both mutagenesis strategies were efficient in their respective host strain; in a typical mutagenesis experiment, an auxotrophy rate of ~1-2% is the norm (G.P.C. Salmond, Pers. comm.). The relative efficiency of mutagenesis in each strain, as indicated by the mean total number of transductants obtained per round of mutagenesis, shows that mutagenesis was more efficient in SCAR1310124 (pMUT13). This could be due to the strain itself, or the transposon used, but since both strains were mutagenised with different transposons, these results are not directly comparable. Mutants differing from wild-type with respect to pigmentation were easily distinguishable (see Fig. 5.3). Table 5.1 shows that the frequencies at which different phenotypes were obtained differed in the two strains. This may be indicative of transposon site-selection, although both Tn*phoA*'-2 and Tn*blaM* carry recombination elements derived from IS*50*L and IS*50*R of Tn*5* (Fig. 5.1).

To exclude the possibility of any of Pig$^-$ mutants being *rap* (regulatory) mutants, further analysis was done (section 5.5). However, it was also considered that some of the non-pigmented transductants might be regulatory mutants in which a positive regulator other than *rap* might have been inactivated by transposon insertion. Also, even if a mutation maps to the *pig* cluster, there is nothing to preclude the possibility of a positive regulator being encoded therein, which might have been affected by transposon insertion.

lacZ gene fusion screening of SCAR1310124::Tn*phoA*'-2 mutants was done after a Lac$^-$ *S. marcescens* strain had been constructed (sections 5.7 and 6.6). SCAR952::Tn*blaM* mutants were screened directly and indirectly for potential fusions (section. 5.6).

5.5.1 CONFIRMATION OF Rap$^+$ STATUS OF NON-PIGMENTED TRANSPOSON MUTANTS

As discussed previously (section 1.3.4.2), a global positive regulator (*rap*) of carbapenem and prodigiosin production in *S. marcescens* has been identified (Thomson *et al.*, 1997). Thomson (1996) had found that approximately

20% of non-pigmented mutants of *S. marcescens* generated by EMS mutagenesis are Rap$^-$. In order to exclude the possibility that some of the Pig$^-$ transposon mutants being defective for this regulator, transductants from each mutagenesis experiment were replicated onto carbapenem bioassay plates as described previously (section 2.9). In addition to this, chromosomal DNAs were prepared from these mutants, as described previously (section 2.3.3), and subjected to PCR using primers specific for *rap*.

5.5.2 CARBAPENEM BIOASSAY RESULTS

A zone of inhibition around the test strain indicated antibiosis due to carbapenem production, because this antibiotic is secreted into the medium where it prevents growth of *E. coli* ESS (McGowan *et al.*, 1996). The size of the inhibitory zone is proportional to the amount of antibiotic produced It was found that all of the Pig$^-$ strains isolated for further study were Car$^+$ (data not shown) and therefore Rap$^+$. Halo sizes on assay plates were the same as for wild type strains SCAR1310124 and SCAR952.

5.5.3.1 PCR ANALYSIS OF PIG$^-$ MUTANTS TO CONFIRM Rap$^+$ STATUS

In order to unequivocally confirm that the Pig$^-$ mutants were not defective for production of Rap, their chromosomal DNAs were subjected to PCR amplification using *S. marcescens rap*-specific primers. Primers to *rap* (Cyto1 and Cyto2, Thomson, 1996, see Appendix II) were provided by N. Thomson and PCR reactions were carried out as described previously (section 2.6),using a primer annealing temperature of 50°C. Following PCR, products were visualised by agarose gel electrophoresis against 1Kb ladder size standards (section 2.3.4.6)

5.5.3.2 RESULTS OF *rap* PCR

Results of agarose gel electrophoresis following PCR are shown for Tn*phoA*'-2 and Tn*blaM* mutants in Fig. 5.3. The predicted product size for *rap*, using Cyto1 and Cyto2 primers is 475bp (Thomson, 1996), and it can be seen that the visualised products following PCR are of this size, and are the same as the

FIGURE 5.3 RESULTS OF PCR TO CONFIRM *rap*$^+$ STATUS OF TRANSPOSON MUTANTS

PCR was carried out using *rap*-specific primers CYTO1 and CYTO2 (Thomson, 1996; Appendix II). The predicted size of the PCR product for wild-type rap is 475bp. Tn*phoA*'-2 insertion in *rap* would result in a product of 8.575 Kb. Tn*blaM* insertion in *rap* would result in a product of 5.725 Kb.

Figure 5.3a

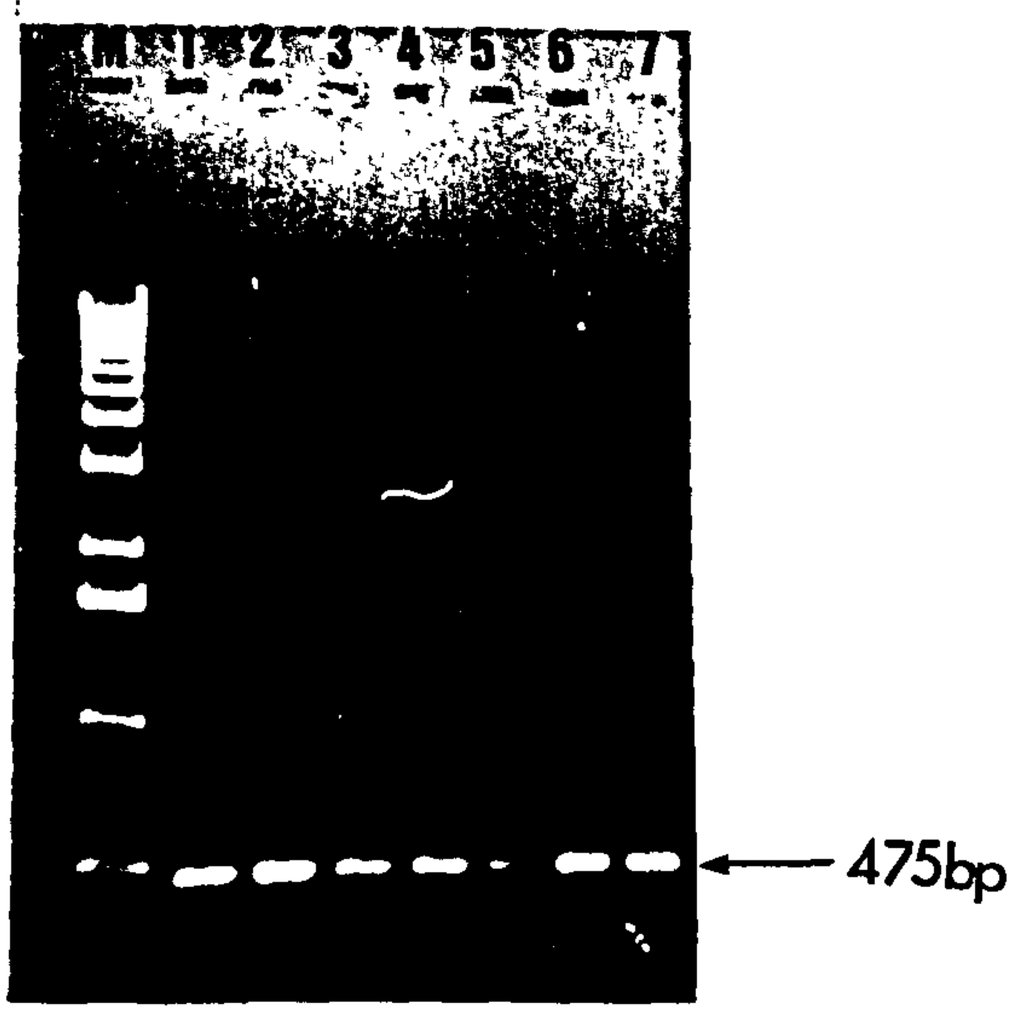


Figure 5.3a Results of PCR on SCAR952::Tn*blaM* mutants
Lanes are numbered and represent the following: M: 1 Kb ladder (Gibco BRL, see section 2.3.4.6); Lanes2-7 represent SCAR952::Tn*blaM* Pig⁻ mutants.
1: SCAR952 wild-type; 2: SYA12; 3: SYA14; 4: SYA20; 5: SYA22;
6: SYA24; 7: SYA25. Control PCR reactions using chromosomal DNAs without primers, and primer without chromosomal DNAs did not result in formation of products (data not shown).

Figure 5.3b

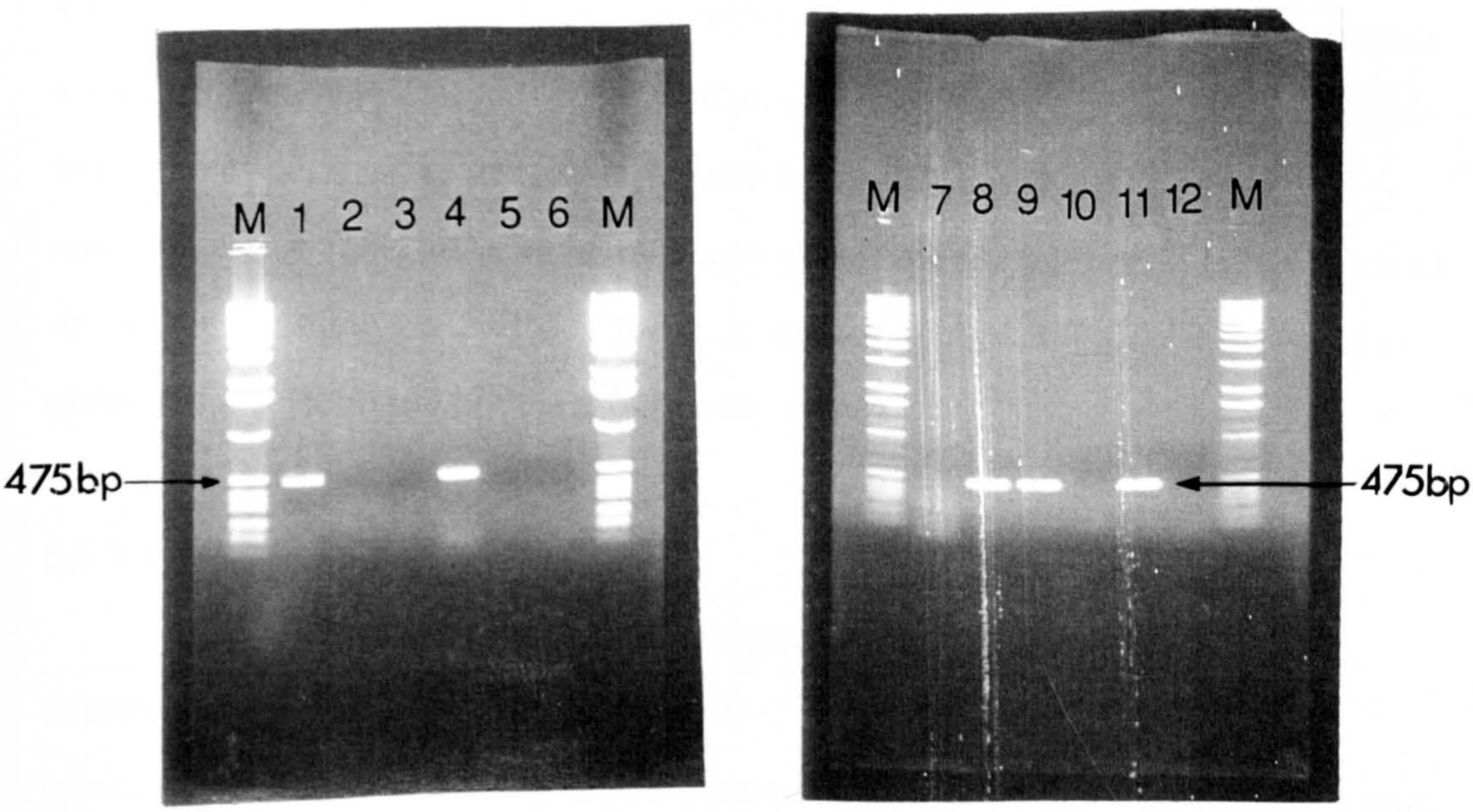


Figure 5.3b Results of PCR on SCAR952::Tn*phoA*'-2 mutants
Lanes are numbered and represent the following: M: 1 Kb ladder (Gibco BRL, see section 2.3.4.6);. 1: SCAR1310124 wild-type; 2: SCAR1310124 DNA, no primers control; 3: SYA100 DNA, no primers control; 4: SYA100; 5: Primers, no chromosomal DNA control; 6: empty lane; 7: SYA101 DNA, no primers control; 8: SYA101; 9: SYA102; 10: SYA102 DNA, no primers control; 11: SYA105; 12: SYA105 DNA, no primers control.

wild-type control. If any of the transposon mutants were defective for *rap*, a band shift corresponding to the size of the inserted transposon would have been seen in comparison with wild-type. It was therefore confirmed that these strains are not *rap* mutants.

5.6.1.1 SCREENING FOR SCAR952::Tn*blaM* TRANSDUCTANTS FOR BlaM ACTIVITY

The *S. marcescens*::Tn*blaM* strains SYA12, SYA14, SYA20, SYA22, SYA24 and SYA25 , SYA30 and SYA50 (Table 2.1) were tested for β-lactamase activity by patching colonies onto NBA containing 5μgml^{-1} of Ap. It was found that none of the strains were Ap^R. In addition to this, another 400 *S. marcescens*::Tn*blaM* mutants which were wild-type for pigmentation were tested for Ap^R in the same way. Out of these, it was found that 24 colonies were able to grow in Ap selection, which is a frequency of 6%.

5.6.1.2 DISCUSSION

To obtain *transcriptional* fusions in transposon mutagenesis experiments, for example, using Tn*phoA*'-2, it would expected that up to a maximum of 50% of resulting transductants would test positive for reporter gene activity, since the only requirement for gene expression is that the reporter gene is inserted in the correct orientation for transcription downstream of a promoter. In experiments where *translational* fusions are to be generated, the chance of obtaining a fusion in the gene of interest is somewhat less than 50% because it is also necessary for the reporter gene to be inserted in-frame within a disrupted host gene, as well as in the correct orientation relative to the promoter. In addition to these constraints, a reporter gene which encodes a protein which is only active in the periplasm, means that the reporter gene need to be fused in-frame in a periplasmic protein or such that the portion of the protein which it encodes in the resultant hybrid polypeptide lies within a periplasmic domain of a cell-membrane integral protein. Because the frequency of obtaining pigment mutants using Tn*blaM* is obviously very low (Table 5.1), it can be argued that results of the fusion screening are a reflection of the small chance by which all the criteria

necessary to obtain a BlaM fusion strain which is also altered for pigment production could be satisfied. The fact that none of the pigment mutants were phenotypically $BlaM^+$ does not necessarily exclude the involvement of periplasmic or transmembrane proteins in prodigiosin biosynthesis in *S. marcescens*, and indeed, as described in Chapter 4, analysis of the encoding region identified three possible putative proteins which might be targeted to the cell membrane. It is possible that *blaM* has inserted in the wrong orientation, or out of frame or in a regions of these genes encoding cytoplasmic domains, and therefore these strains are phenotypically $BlaM^-$.

Tn*blaM* mutants which were able to grow at high colony density when patched onto NBA-Ap were presumed to be BlaM fusion strains, as high density growth indicates that fusions to extracytoplasmic regions of host-encoded proteins were obtained. The frequency at which this occurred (6%) is much less than that which has been previously reported for *E. coli*, which was 30% (Taddyon and Broome-Smith, 1992). This might be due to differences in the efficiency of Tn*blaM* transposition in these species.

5.6.2.1 SCREENING FOR BlaM FUSIONS USING SECONDARY TRANSPOSITION SELECTION

In addition to transductant colonies which grew at high density on NBA-Ap, further experiments were done to try and identify any transductant strains which are efficient for Tn*blaM* transposition, to see if selection on NBA-Ap results in the production of any $BlaM^+$ mutants which are also altered for pigment production. This selection procedure for secondary transposition events, has been used successfully in *E. coli* (Taddyon and Broome-Smith, 1992). Tn*blaM* mutants which are unable to grow on NBA-Ap at single-cell density, but which have cytoplasmic BlaM fusions can yield Ap^R colonies when patched onto NBA-Ap at concentrations exceeding their single-cell minimum inhibitory concentrations (MICs). The appearance of survivors in patches which can then be maintained on high-concentration Ap-agar is indicative of secondary transposition events where *blaM* has fused in frame to a portion of a gene encoding an extracytoplasmic domain. Patch testing of 500 non-Ap^R colonies on NBA-Ap

($10\mu g\ ml^{-1}$), which were grown at 30°C for four days did not yield any non-pigmented Ap^R colonies.

5.6.2.2 DISCUSSION OF SECONDARY TRANSPOSITION SELECTION EXPERIMENTS

The efficiency with which Tn*5* and its derivatives transpose is influenced by the amount of transcription running into the ends of the transposon from flanking chromosomal promoters (Sasakawa *et al.*, 1982) and therefore, the efficiency of secondary transposition depends upon precise chromosomal location of the transposon. From these results, it appears that strain SCAR952 is not particularly susceptible to secondary transposition events, and this might reflect low levels of transcription of flanking chromosomal genes. It is possible that the transposon is hot-spotting into certain areas, although Tn*5* has relatively low site-selection specificity (Berg, 1977; Shaw and Berg, 1979; Berg *et al.*, 1980; Miller, *et al.*, 1980; Berg, 1983). When the earlier phenotype testing results (section 5.5.1) are considered, it seems quite improbable that use of secondary transposition screening would have resulted in the production of a $BlaM^+$ mutant which is also affected in pigment phenotype. However, since the use of this transposon, together with the strategy for selecting for secondary transposition constitutes a technically facile method for generating fusions, pursuing this line of experimentation was believed to be justified at the time.

As T*nblaM* mutagenesis did not result in the construction of phenotypically $BlaM^+$ strains and selection for secondary transposition failed to produce any relevant fusion strains, no further work was done on these mutants; once it was established that Tn*phoA*'-2 mutagenesis had been successful in generating fusion strains (described in Chapter 6) only they were analysed further. Tn*blaM* mutants which were generated in this study could be, in the future, characterised further in a similar way that Tn*phoA*'-2 mutants were in this study (see Chapter 6), but time constraints meant that they were not investigated further.

5.7.1 CONSTRUCTION AND MUTAGENESIS OF A LAC⁻ STRAIN OF *Serratia marcescens*

As discussed in section 5.4, Tn*phoA*'-2 pigment mutants generated in strain SCAR130124 could not be screened for potential *lacZ* fusions to pigment genes because the strain is Lac^+. The problem was overcome by the construction of a Lac^- strain of *S. marcescens*, and the use of the *S. marcescens* generalised transducing phage ΦOT8 to transduce mutations over to the Lac^- background (described in Chapter 6). Additionally, once a Lac^- strain was obtained, it was also thought worthwhile to also mutagenise the strain using λTn*phoA*'-2. By this approach, more gene fusion strains were obtained directly.

5.7.2.1 EMS MUTAGENESIS OF SCAR1310124

Strain SCAR1310124 was mutagenised using ethyl methane sulphonate (EMS, formula $CH_3SO_3CH_2CH_3$, Maloy *et al.*, 1994b). This chemical mutagen is an alkylating agent, which exerts its mutagenic action by the addition of an alkyl group to the hydrogen-bonding oxygen of guanine and thymine. Impairment of normal hydrogen-bonding causes in mispairing of G and T, resulting in GC→AT transitions and vice versa, with the former being predominant (Maloy *et al.*, 1994b). This was the method of choice because it is efficient and does not involve the introduction of additional genetic information into the chromosome.

5.7.2.2 RESULTS OF EMS MUTAGENESIS

The protocol used is described in section 2.8. The killing curve which was constructed (data not shown), showed that 1% survival occurs at 130 min for SCAR1310124. The experiment was repeated with a 130 min EMS exposure time and the culture was serially diluted and plated out as described previously (section 2.8). The plating medium used was NBA-X-Gal, in order to directly screen for Lac^- mutants, which appear as white colonies, compared with Lac^+ colonies which appear blue. Prodigiosin production by *S. marcescens* would have complicated the visual screening of white versus blue colonies, so plates were therefore

incubated at 37°C for 4-5 days to abolish prodigiosin production and any white colonies were retained for further analysis. It was found that out of a total 7,300 colonies screened, 3 appeared white on NBA-X-Gal, which is a frequency of 0.04%. An additional 370 colonies were randomly picked and replicated onto NBA and MMA and incubated at 30°C for 3 days. This revealed an auxotrophy rate of 1.4% for the mutagenesis experiment. Putative Lac$^-$ colonies were replicated onto NBA-X-Gal and incubated at 37°C and also NBA and MMA and incubated at 30°C. This resulted in the isolation of strain SLAC1 which is prototrophic and Pig$^+$.

5.7.2.3 FURTHER PHENOTYPE CONFIRMATIONS

Before the strain was used for molecular genetic and physiological studies, a few experiments were done to ensure that the strain is not pleiotropically mutant as far as could be ascertained. The strain was confirmed to be Car$^+$ on the basis of the carbapenem bioassay using *E .coli* ESS (section 2.9) with SCAR1310124 as positive control. It was found to be susceptible to transduction by phage ΦOT8 on the basis of a phage titration assay performed as described in sections 2.5.3 and 2.5.1. This showed that, by the appearance of discrete plaques, ΦOT8 is able to lyse the cells from within, excluding the possibility of lysis from without.

In order to confirm the Lac$^-$ status of this strain, a 10 ml overnight was grown in LB at 30°C. A 1 ml sample of this was subjected to cell disruption by toluene treatment (section 2.12.1). Aliquots of 10µl, 50µl, 100µl, and 250µl of this were added to Z buffer to a final volume of 500µl for each sample. LacZ assays were performed on these samples, as described in section 2.12.2. It was found that LacZ activity was negligible in these samples and the colour change normally indicative of LacZ activity was barely detectable above the blank even after a 2h incubation. The strain was then used for further genetic analyses, including the screening of SCAR1310124::Tn*phoA*'-2 pigment mutants for LacZ fusions, but in order to expand the repertoire of pigment mutant fusion strains, SLAC1 was firstly subjected to λTn*phoA*'-2 mutagenesis.

5.7.3.1 λTn*phoA*'-2 MUTAGENESIS OF SLAC1

Prior to mutagenesis it needed to be established if SLAC1 had inherited functional pMUT13 from the SCAR1310124 progenitor strain. Serial dilution and plating on NBA-Kn that SLAC1 showed that SLAC1 was unable to grow on this medium, so it was assumed that pMUT13 had been lost (or mutated) in the course of EMS mutagenesis or subsequent culturing, in which Kn selection had not been maintained. pMUT13 was provided by J. Thomas. SLAC1 was transformed with this plasmid by electroporation (section 2.4.3) and transformants were selected on NBA-Kn.

Several overnight cultures of SLAC1(pMUT13) were transduced using a high titre λTn*phoA*'-2 lysate as described previously (section 2.5.2). Transductants were selected on NBA-Tc-X-Gal and plates were incubated at 37°C for 4-5 days. Incubation at this temperature abolished pigment production so that blue-white selection could be done. Blue colonies were replicated onto NBA-X-Gal (incubated at 37°C), NBA and MMA and incubated at 30°C. Colonies which were Pig^- and prototrophic at 30°C were isolated for further study. A further 300 transductant colonies were randomly picked for each experiment and replicated onto NBA and MMA and incubated for 3 days at 30°C in order to assess auxotrophy frequency.

5.7.3.2 RESULTS OF λTn*phoA*'-2 MUTAGENESIS OF SLAC1

Results of mutagenesis experiments are presented in Table 5.2. Table 5.2 shows that results of the mutagenesis experiment are very similar in terms of phenotypes obtained and their frequencies to those of λTn*phoA*'-2 mutagenesis of the progenitor strain SCAR1310124. Visual analysis of NBA-X-Gal-Tc plates revealed that on average, approximately 44% of transductants were LacZ fusions. From these experiments, 2 strains which were Lac^+, Pig^- and protrophic were isolated for further study (Table 5.2).

TABLE 5.2 RESULTS OF λTn*phoA*'-2 MUTAGENESIS OF SLAC1

MUTANT PHENOTYPE	% MEAN FREQUENCY PER TRANSDUCTION[a]
Non-pigmented (Pig^-)	0.28
Hyperpigmented (Pig^{++})	0.1
Orange (Pig^O)	7.3
Auxotrophic	1.4

[a]Frequencies were derived by comparing the number of colonies of each mutant phenotype with the total number of colonies screened. Mean total number of colonies per transduction was 980

5.7.4 CONFIRMATION OF RAP^+ STATUS OF SLAC1::Tn*phoA*'-2 Pig^- MUTANTS

Results of the carbapenem bioassay (Section 2.9, data not shown) and PCR using *rap* primers Cyto1 and Cyto2 as described previously, (section 5.5.4.1) showed that the Pig^- SLAC1::Tn*phoA*'-2 strains which were retained for further study were Rap^+ (Fig. 5.4)

5.8 GENERAL DISCUSSION AND SUMMARY

In experiments with both types of transposon, it was found that orange mutants were obtained at significantly higher frequencies than other pigment phenotypes. It is possible that these mutants are of the kind reported previously (Morrison, 1966) which are blocked in the methylation of MBC so produce norprodigiosin in place of prodigiosin (see section 1.3.3). The increased frequency at which the orange phenotype is obtained might again be indicative of transposon site-selection bias.

The hyperpigmented transductants which were obtained were presumed to be regulatory mutants, since it is the level of pigment gene expression which is affected. The most obvious and simple possible reason for is that a repressor gene has been insertionally inactivated in these mutants. The lower frequency at which these mutants were obtained compared with orange (presumably biosynthetic) mutants and non-pigmented (which were thought most likely to be biosynthetic knock-out) mutants provides a clue to the genetics of pigment production, because these results might be indicative of there being more

Figure 5.4

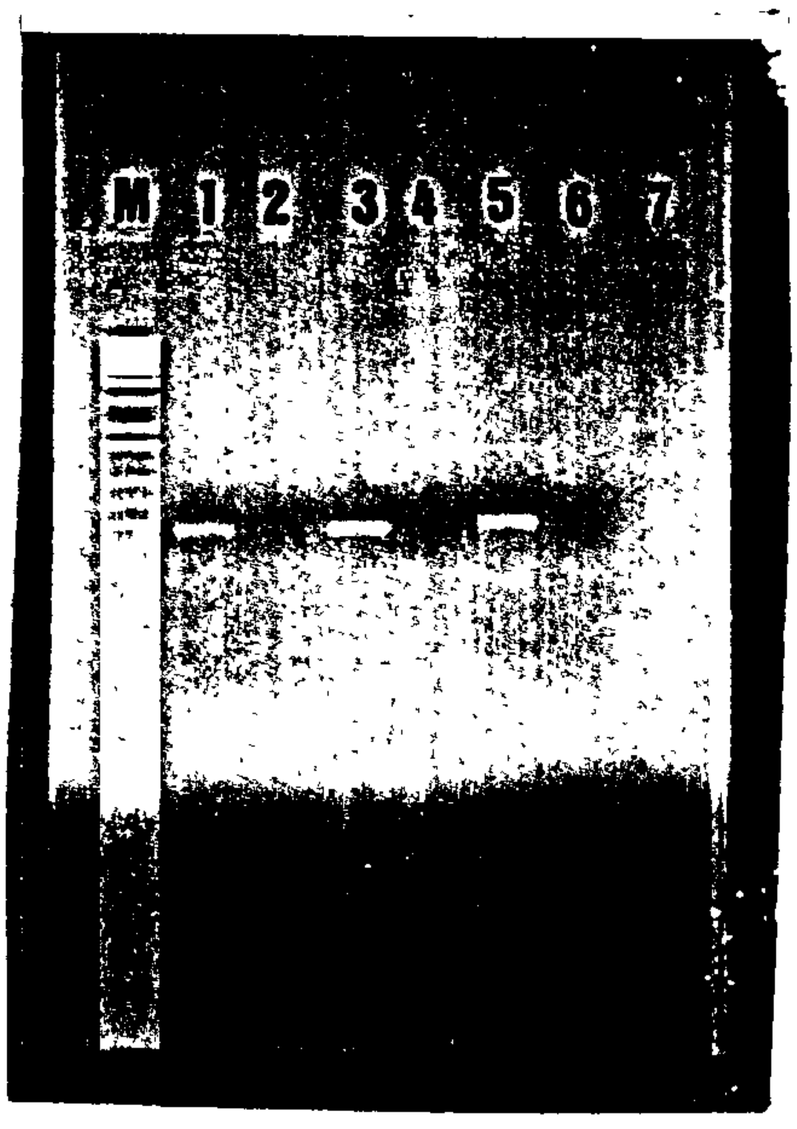


FIGURE 5.4 RESULTS OF PCR TO CONFIRM *rap*$^+$ STATUS OF SLAC1::Tn*phoA*'-2 MUTANTS

PCR was carried out using *rap*-specific primers CYTO1 and CYTO2 (Thomson, 1996; Appendix II). The predicted size of the PCR product for wild-type rap is 475bp. Tn*phoA*'-2 insertion in *rap* would result in a product of 8.575 Kb. Lanes are numbered and represent the following: M: 1 Kb ladder (Gibco BRL, see section 2.3.4.6);. 1: SLAC1 wild-type; 2: SLAC1 DNA, no primers control; 3: SYA111 DNA, 4: SYA111 DNA , no primers control; 5: SYA113; 6: SYA113 DNA, no primers control; 7: Primers, no chromosomal DNA control.

Figure 5.4

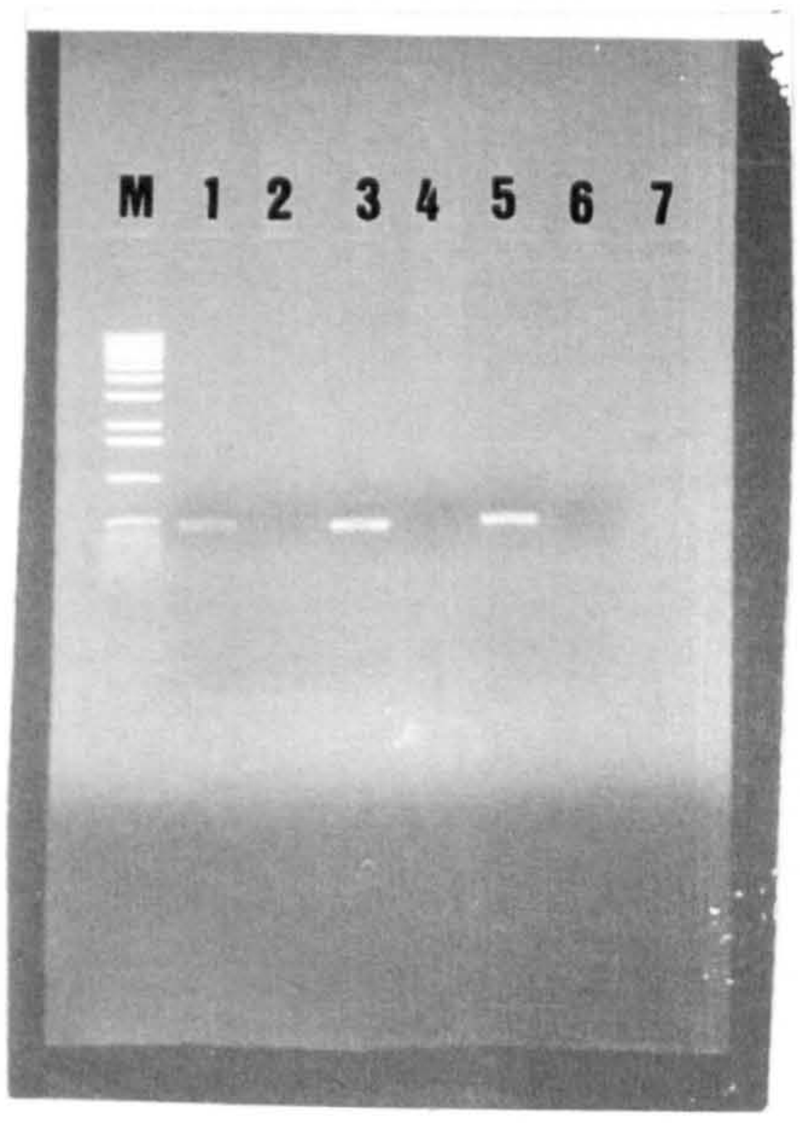


FIGURE 5.4 RESULTS OF PCR TO CONFIRM *rap*$^+$ STATUS OF SLAC1::Tn*phoA*'-2 MUTANTS

PCR was carried out using *rap*-specific primers CYTO1 and CYTO2 (Thomson, 1996; Appendix II). The predicted size of the PCR product for wild-type rap is 475bp. Tn*phoA*'-2 insertion in *rap* would result in a product of 8.575 Kb. Lanes are numbered and represent the following: M: 1 Kb ladder (Gibco BRL, see section 2.3.4.6);. 1: SLAC1 wild-type; 2: SLAC1 DNA, no primers control; 3: SYA111 DNA, 4: SYA111 DNA , no primers control; 5: SYA113; 6: SYA113 DNA, no primers control; 7: Primers, no chromosomal DNA control.

prodigiosin biosynthetic genes present on the *S. marcescens* chromosome than regulatory genes which control the level of pigment biosynthetic gene expression.

In summary, transposon mutagenesis was successfully used to mutagenise three *S. marcescens* strains. The mutant strains, which were altered in their pigment phenotypes with respect to wild-types, were tagged with transposon insertions, and this was used as the basis for molecular characterisation of some of theses strains. In subsequent work, it was decided to concentrate on any mutants in which the transposon insertions could be mapped to the biosynthetic cluster. Although extremely interesting, further investigation of potential regulatory mutants was not within the scope of this project. Molecular genetic analysis work which was done to characterise Tn*phoA*'-2 mutants is presented in Chapter 6.

CHAPTER 6

FURTHER CHARACTERISATION OF Tn*phoA*'-2 PIGMENT MUTANTS

6.1 PREFACE

Strains generated by Tn*phoA*'-2 mutagenesis of SCAR1310124 (pMUT13) and SLAC1 (pMUT13) were characterised further by Southern blotting. The presence of the transposon in each mutant strain was exploited in cloning of inactivated genes. The work also confirmed that some useful gene fusion strains had been obtained which allowed some expression studies to be undertaken in the final part of the study. Some of these fusion strains were constructed by generalised transduction experiments which transferred Tn*phoA*'-2 mutations from a Lac^+ to a Lac^- background. Localisation of mutations in different strains to the *pig* cluster would provide further evidence to support data from the sequencing project.

6.2 AIMS

The overall aim of the work presented in this chapter was to use transposon insertions in Tn*phoA*'-2 mutants to identify genes required for prodigiosin biosynthesis which had been inactivated by transposon mutagenesis.

6.3.1 PROBING Tn*phoA*'-2 MUTANTS WITH A TRANSPOSON SPECIFIC PROBE

In order to obtain physical evidence of transposon insertions in the Pig^- mutants, it was decided to probe chromosomal DNAs from these strains with a transposon-specific probe. The probe used is 300bp long and was generated by digoxigenin-11-dUTP (DIG) labelling by PCR (Boehringer Mannheim); it hybridises equally to both ends of Tn*phoA*'-2 (S. Harris, Pers. comm.). The probe was provided by M. Chan.

Chromosomal DNAs were prepared from SCAR1310124 and SLAC1 wild-types and the non-pigmented strains SYA100, SYA101, SYA102, SYA105, SYA111 and SYA113. It was found during the course of this study that *S. marcescens* chromosomal DNA was difficult to prepare. Although the preparation method (described in section 2.3.3), was successful in lysing cells, obtaining sufficiently high concentrations of DNA for chromosomal digestion and Southern blotting proved problematic, as it was found that DNA was difficult to resuspend

after ethanol precipitation. It was also found that chromosomal DNA from Pig^{++} and Pig^{O} mutants was particularly difficult to prepare and as a result, attempts at probing DNA from these mutants resulted in failure to detect signal from DIG labelled probes. This meant that the investigation was then focused on characterisation of Pig^{-} mutants for the remainder of the study.

Chromosomal DNAs were digested using *Eco*RI, as described previously (section 2.3.4.1). This enzyme was chosen because Tn*phoA*'-2 has a unique *Eco*RI restriction site at nt5460 (see Fig. 5.1), so two bands, per transposon insertion, would be visualised in mutants after probing. Southern blot transfer of DNA to Hybond-N™ nylon membrane was carried out as described in section 2.7 and the filter was probed overnight at 55°C. Results were visualised by autoradiography after a 4h time exposure.

6.3.2 RESULTS AND DISCUSSION OF TRANSPOSON PROBING

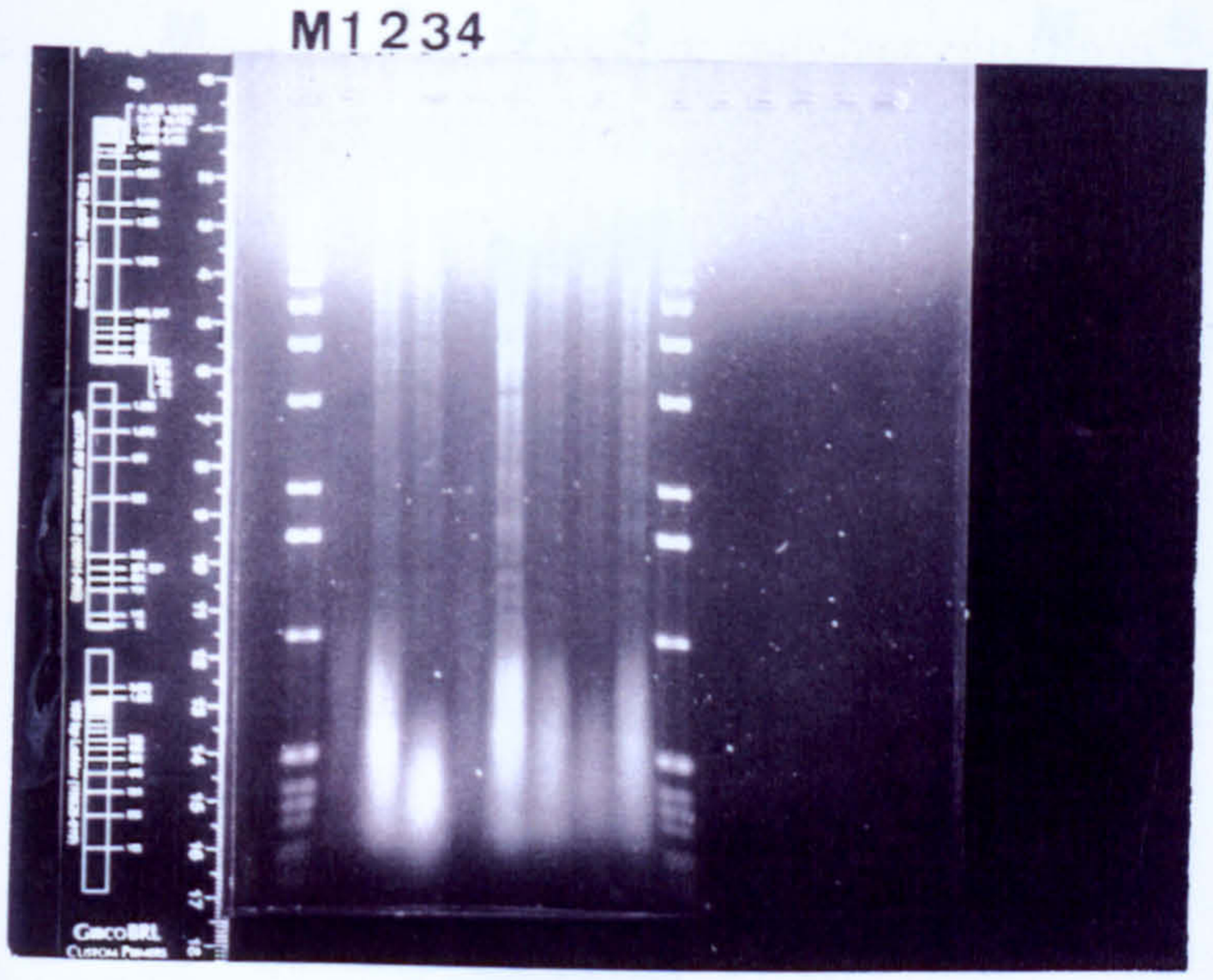
Results of Southern blotting are presented in Fig. 6.1a and Fig 6.1b. Fig. 6.1b shows that only two bands are present per lane, indicating that single transposon insertions are present in each mutant. Probing of both SCAR1310124 and SLAC1 did not show any bands (data not shown), as would be predicted for wild-type, non-mutagenised strains. The presence of single transposon insertions in SCAR1310124::Tn*phoA*'-2 mutants was also confirmed later by generalised transduction (discussed in section 6.6).

The sizes of the bands produced in these experiments depends on the insertion site of the transposon and for any mutant, band sizes are determined by the location of the nearest *Eco*RI sites in the flanking chromosomal DNA. Cloning of transposon insertions (described in section 6.4) identified the approximate insertion site in each mutant. Results of transposon probing therefore were re-analysed after cloning. As Southern blotting showed that single insertions are present in each mutant strain tested, experiments to clone the insertions could then proceed.

FIGURE 6.1 RESULTS OF PROBING *Serratia marcescens*::Tn*phoA*'-2 Pig⁻ MUTANTS WITH A TRANSPOSON-SPECIFIC PROBE

Chromosomal digests and corresponding Southern blots are shown in Fig.6.1a and Fig. 6.1b respectively. *Eco*RI digested chromosomal DNAs were probed with a DIG-labelled TnphoA'-2 specific probe. See text in sections 6.3.2 and 6.4.4.1 for details and discussion of results

Figure 6.1a



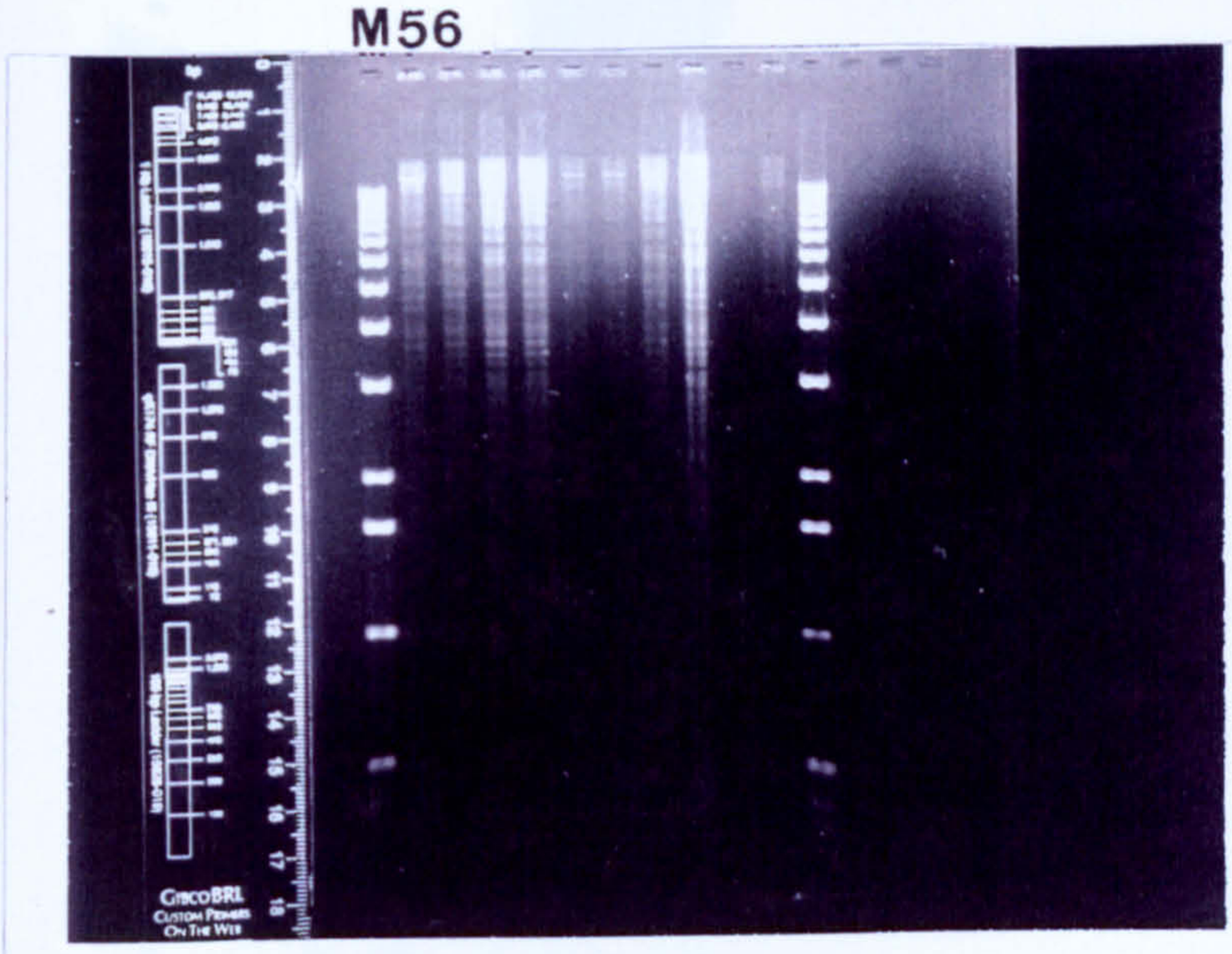


Figure 6.1a *Eco*RI digests of chromosomal DNAs.
Key: M: 1 Kb ladder (Gibco BRL, see section 2.3.4.6); 1: SYA100; 2: SYA101; 3: SYA102; 4: 105; 5: SYA113; 6: SYA111.

Figure 6.1b

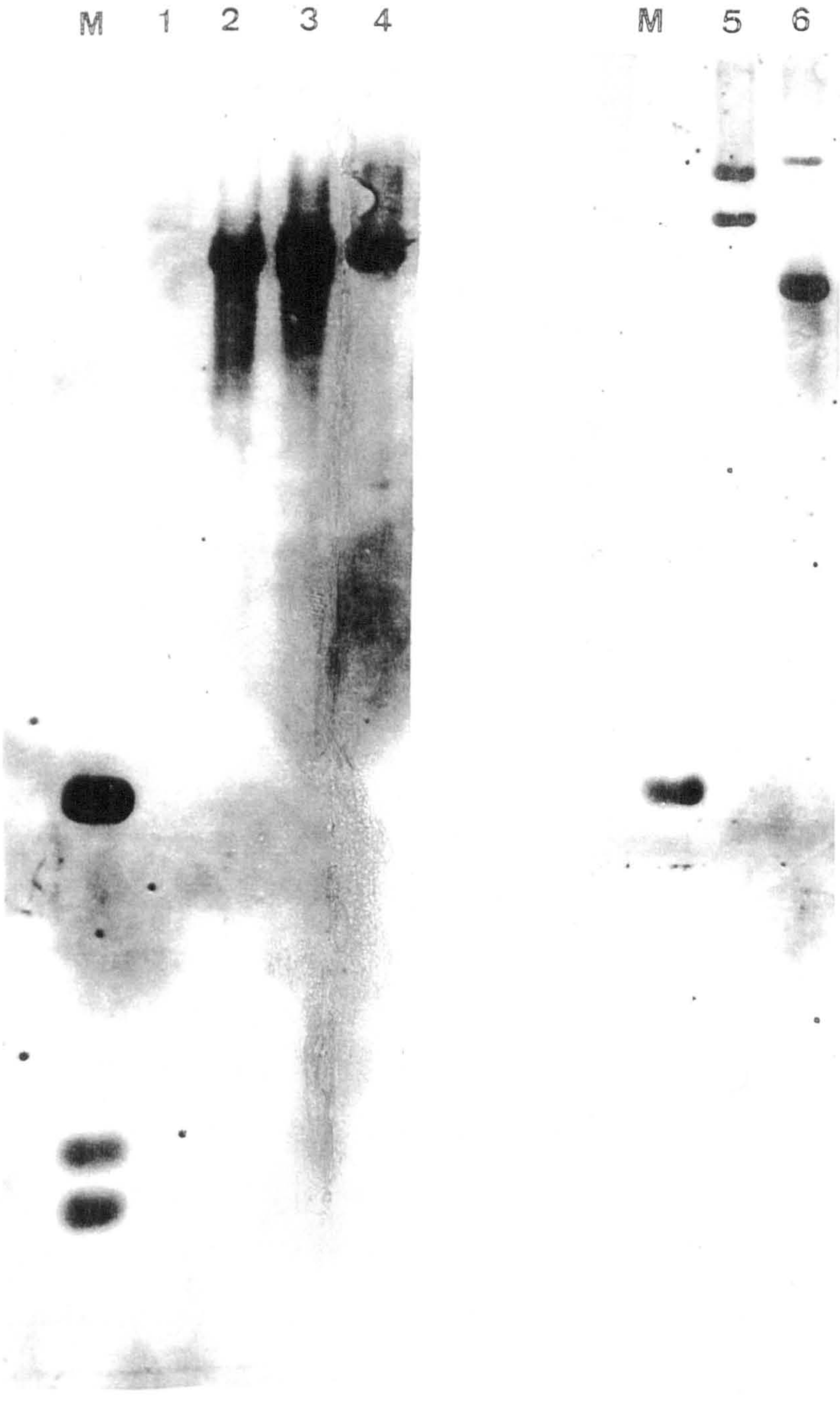


Figure 6.1b Southern blots of Tn*phoA*'-2 Pig⁻ mutants
Key: M: 1Kb Ladder (Gibco BRL, see section 2.3.4.6); 1: SYA100; 2: SYA!01 3: SYA102; 4: SYA105; 5: SYA113; 6: SYA111. The signal detected in lane M is caused by non-specific binding of this probe to 1Kb ladder. This phenomenon is usually observed with this particular probe (S. Harris, M. Chan, Pers. comm.)

6.4.1 CLONING OF Tn*phoA*'-2 INSERTIONS FROM Pig⁻ MUTANTS

Tn*phoA*'-2 mutants and one side of flanking chromosomal DNA were cloned to localise insertion sites in the *S. marcescens* chromosome. This cloning strategy relies upon the presence of a unique restriction sites within the transposon sequence, as well as the presence of a selectable marker (Tc^R in Tn*phoA*'-2). In this study, DNA flanking IS*50*R was cloned out on a *Bam*HI fragment and sequenced to identify inactivated loci.

The cloning strategy is summarised in Fig. 6.2. Pig⁻ strains SYA100, SYA101, SYA102, SYA105, SYA111 and SYA113 were cloned as follows: chromosomal DNAs were digested with *Bam*HI in a total volume of 40µl as described previously (section 2.3.4.1). DNA was run out on an agarose gel against 1Kb Ladder size standards and DNA of size >7Kb was excised from the gel and extracted using GENECLEAN II™ (section 2.3.4.7). DNA was ligated into the low copy vector pACYC177 (Appendix I), which encodes Ap^R and Kn^R, and had been digested with *Bam*HI and treated with calf alkaline phosphatase as described previously (section 2.3.4.2). Cloning of Tn*pho*A'-2 requires the use of a low copy vector because the level of Tc^R conferred on recombinant clones is greatly reduced if the transposon is present in high copy. *Bam*HI fragments of chromosomal DNA were ligated to vector as described previously (section 2.3.4.4) and transformed into *E. coli* DH5α by $CaCl_2$ transformation (section 2.4.1). Transformants were selected on NBA-Kn-Tc. Several clones from each experiment were picked, grown in NB with selection and plasmid DNA prepared from then by alkaline lysis (section 2.3.1). *Bam*HI digestion and agarose gel electrophoresis verified that clones contained single inserts >7Kb in length. One clone from each experiment was then used to prepare high-grade plasmid DNA by Midi-Prep (Qiagen, section 2.3.2) and DNA was submitted for in-house automated sequencing (section 2.8.2). Tn*phoA*'-2 Primer 1 (Appendix II, provided by M. Sebaihia) was used for sequencing reactions. This primer anneals to the extreme end of IS*50*R in these clones, 13nt from the end of the transposon.

Figure 6.2

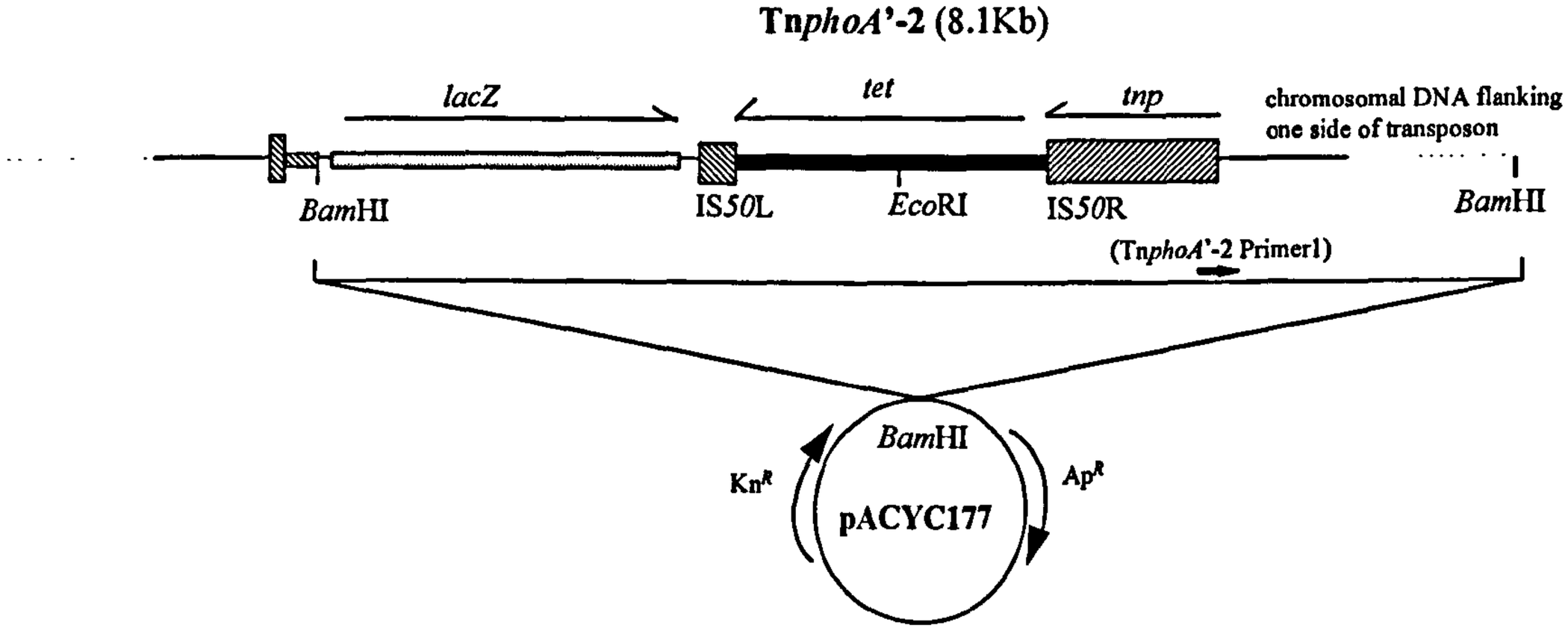


FIGURE 6.2 STRATEGY USED TO CLONE Tn*phoA*'-2 INSERTIONS FROM THE CHROMOSOMES OF Pig MUTANTS

The size of inserted DNA in the recombinant clone depends upon the location of the nearest *Bam*HI site in the chromosome. *Bam*HI digestion of chromosomal DNA carrying Tn*phoA*'-2 does not destroy Tc^R. See text in 6.4.1 for details.

6.4.2 RESULTS OF CLONING AND SEQUENCING OF Tn*phoA*'-2 INSERTIONS

Cloning of transposon DNA between its *Bam*HI site and IS*50*R, and flanking chromosomal DNA was successful for the Pig$^-$ strains. Sequencing data from each clone were firstly compared with both strands of the contiguous sequence of the *pig* cluster and aligned wherever possible. Secondly, sequences were submitted into searches of non-redundant databases on the WWW using BLAST programs (Appendix III). Results of these analyses are summarised in Table 6.1. Recombinant plasmids carrying Tn*phoA*'-2 insertions from each Pig$^-$ strain are listed in Table 2.2

TABLE 6.1 SUMMARY OF FINDINGS OF TRANSPOSON CLONE SEQUENCE ALIGNMENTS WITH THE *pig* CLUSTER AND DATABASE SEARCHES

STRAIN AND PLASMID CONSTRUCT	MOST SIMILAR DATABASE HOMOLOGUE(S) FOUND IN BLAST-X SEARCH	CODING OR NON-CODING STRAND	DEDUCED RANGE OF INSERTION POINT IN *pig* CLUSTER NUCLEOTIDE SEQUENCE[a] (nt)
SYA100 pACYC177/100	OmpR (*Sal. typhimurium*) EnvZ (*Y. enterocolitica*)	- -	none
SYA101 pACYC177/101	RedW (*Str. coelicolor* A3(2))	+	2281-2308
SYA102 pACYC177/102	Probable PEP-utilising enzyme (*Str. coelicolor* A3(2))	-	6800-6833
SYA105 pACYC177/105	HscA (*E. coli*)	-	none
SYA111 pACYC177/111	none	-	8853-8883
SYA113 pACYC177/113	RedW (*Str. coelicolor* A3(2))	+	1680-1710

[a]: The insertion point of Tn*phoA*'-2 in each strain, where relevant, was deduced by sequence alignment with the contiguous sequence of the *pig* cluster. Nucleotide numbers correspond to numbering of the contiguous sequence (refer to Fig. 3.4, Chapter 3 and Appendix IV). See also figure 6.3b.

On the basis of sequence data, the transposon insertion point in each strain was narrowed down to within a range of 30nt around the precise probable insertion point. Sequences from four constructs were found to map to the *pig* cluster (Table 6.1 and Fig. 6.3)

The remaining two clones, from SYA100 and SYA105 did not align with the *pig* cluster DNA on either strand. It was therefore assumed that in these mutants, genes outside of the *pig* cluster, which are essential for prodigiosin biosynthesis, or expression of prodigiosin biosynthetic genes, had been insertionally inactivated. Further evidence of this was obtained when mutants were probed with *pig* cluster DNA (discussed in section 6.5.1). Database searches provided further evidence to confirm what sequences alignments had shown (see Table 6.1). Plasmid pACYC177/111, derived from SYA111 does not have significant homology to any known database proteins, and the same was found to be true of *orf4*, to which this transposon insertion maps in the *pig* cluster.

Sequences derived from SYA100 and SYA105 were found to be highly similar to known database proteins. The sequence from SYA100 is homologous to both OmpR and EnvZ from various prokaryotes. The translated sequence of pACYC177/100, corresponding to the latter half of the actual nucleotide sequence, was found to share 92% identity and 93% similarity over a 160 amino acid overlap with OmpR of *Salmonella typhimurium*. This region was designated *orfX*. The deduced amino acid sequence corresponding to the first half of the nucleotide sequence of pACYC177/100 shares 67% identity and 69% similarity with EnvZ of *Yersinia enterocolitica*, in an 85 residue overlap, and this region was designated *orfY*. Database searches using sequence derived from SYA105 transposon clone pACYC177/105 showed that it is homologous to *hscA* of *E. coli* and *Haemophilus influenzae*. The translated sequence shares 81% identity and 86% similarity with *E. coli* HscA over a 106 amino acid overlap. This region was designated *orfZ*. Regions of translated sequences from SYA100 and SYA105 which are homologous to the database proteins are depicted in Fig. 6.4.

Figure 6.3a

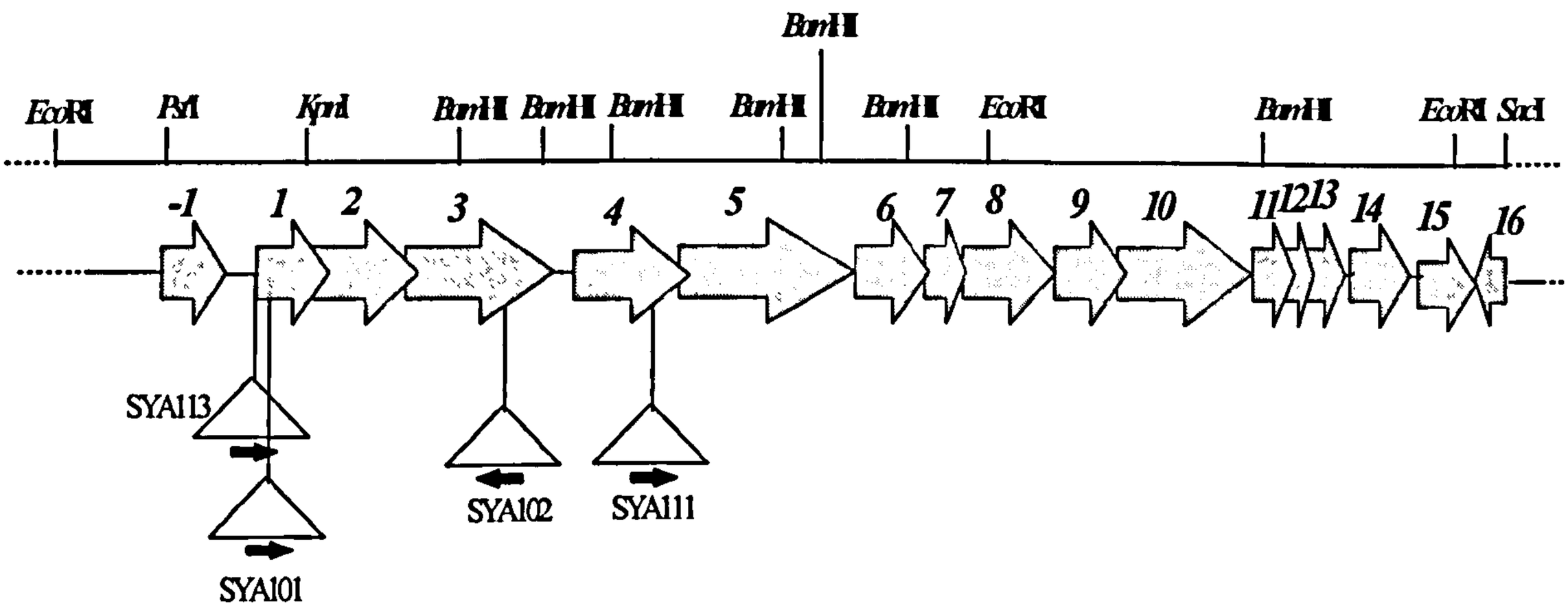


FIGURE 6.3a DEDUCED APPROXIMATE INSERTION POINTS OF Tn*phoA*'-2 IN THE *pig* CLUSTER OF DIFFERENT Pig$^-$ STRAINS

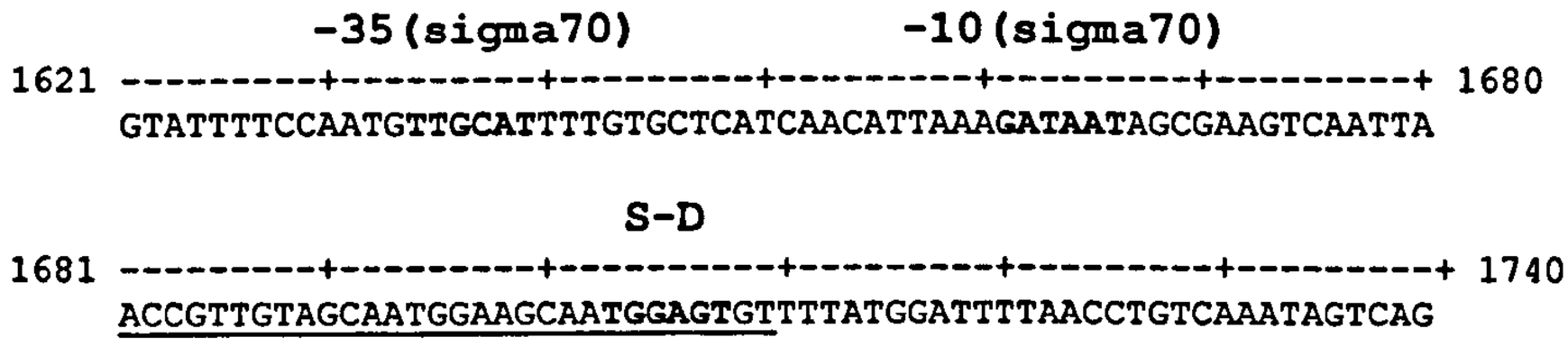
Large grey arrows represent designated ORFs in the *pig* cluster, numbered accordingly. Shown above this is a restriction map of relevant enzyme sites, as deduced from the prodigiosin biosynthetic cosmid pNRT104 and plasmid pM245. Black arrows below Tn*phoA*'-2 insertions represent transcriptional orientation of *lacZ*, and hence the transposon, with respect to the disrupted ORF. See text in section 6.3.2 for details.

FIGURE 6.3b REGIONS IN WHICH Tn*phoA*'-2 INSERTIONS ARE PRESENT IN THE *pig* CLUSTER

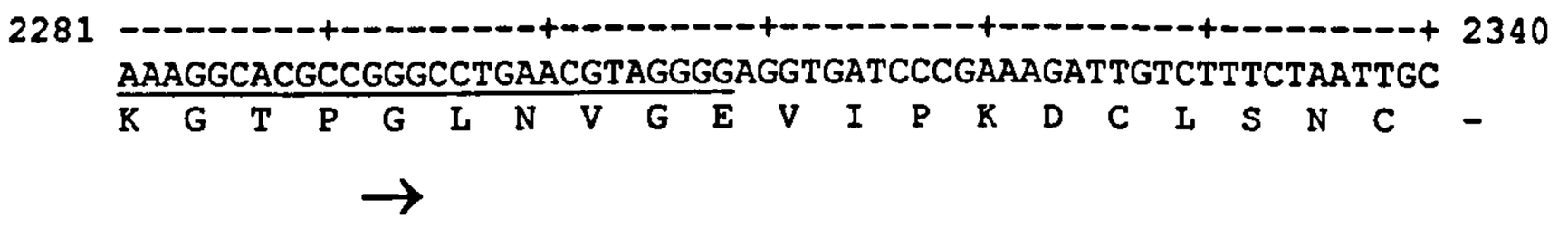
For each strain, the DNA region in which the Tn*phoA*'-2 insertion is believed to be present is indicated by underlining. The exact insertion points of transposons are not known, but it was possible to limit the exact insertion point to a narrow region by nucleotide sequence alignments. Precise pinpointing of insertion sites would require more sequencing. For clarity, only the non-coding strand, and translated amino acid sequence, where relevant, is shown in each case. Numbers next to coding strand correspond to nucleotides in the *pig* cluster and are the same as the contiguous sequence presented in Figure 3.4, and Appendix IV. The arrow below each underlined region represents the orientation of Tn*phoA*'-2 in each strain. Also shown in **bold**, where relevant, are possible transcriptional motifs which were identified by sequence analysis (see Chapter 3).

Figure 6.3b

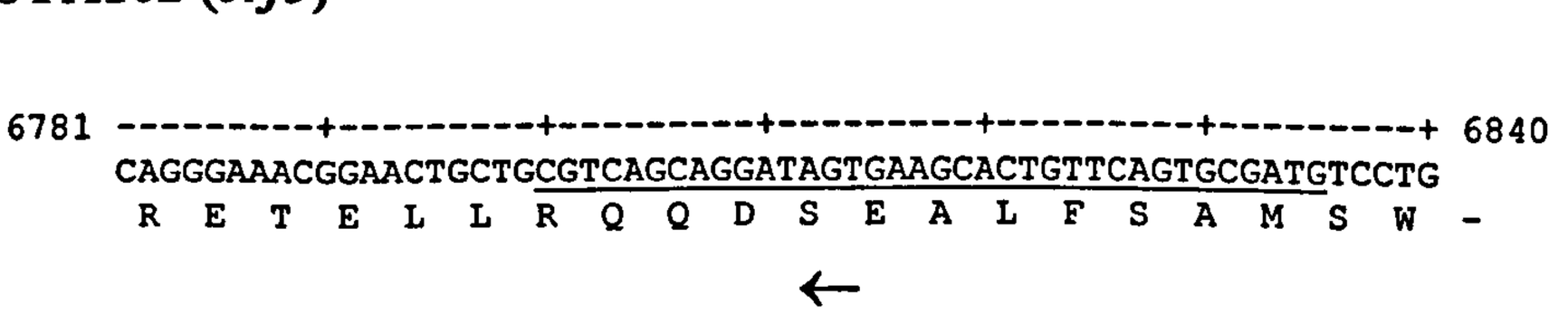
(i) SYA113



(ii) SYA101 (*orf1*)



(iii) SYA102 (*orf3*)



(iv) SYA111 (*orf4*)

Figure 6.4

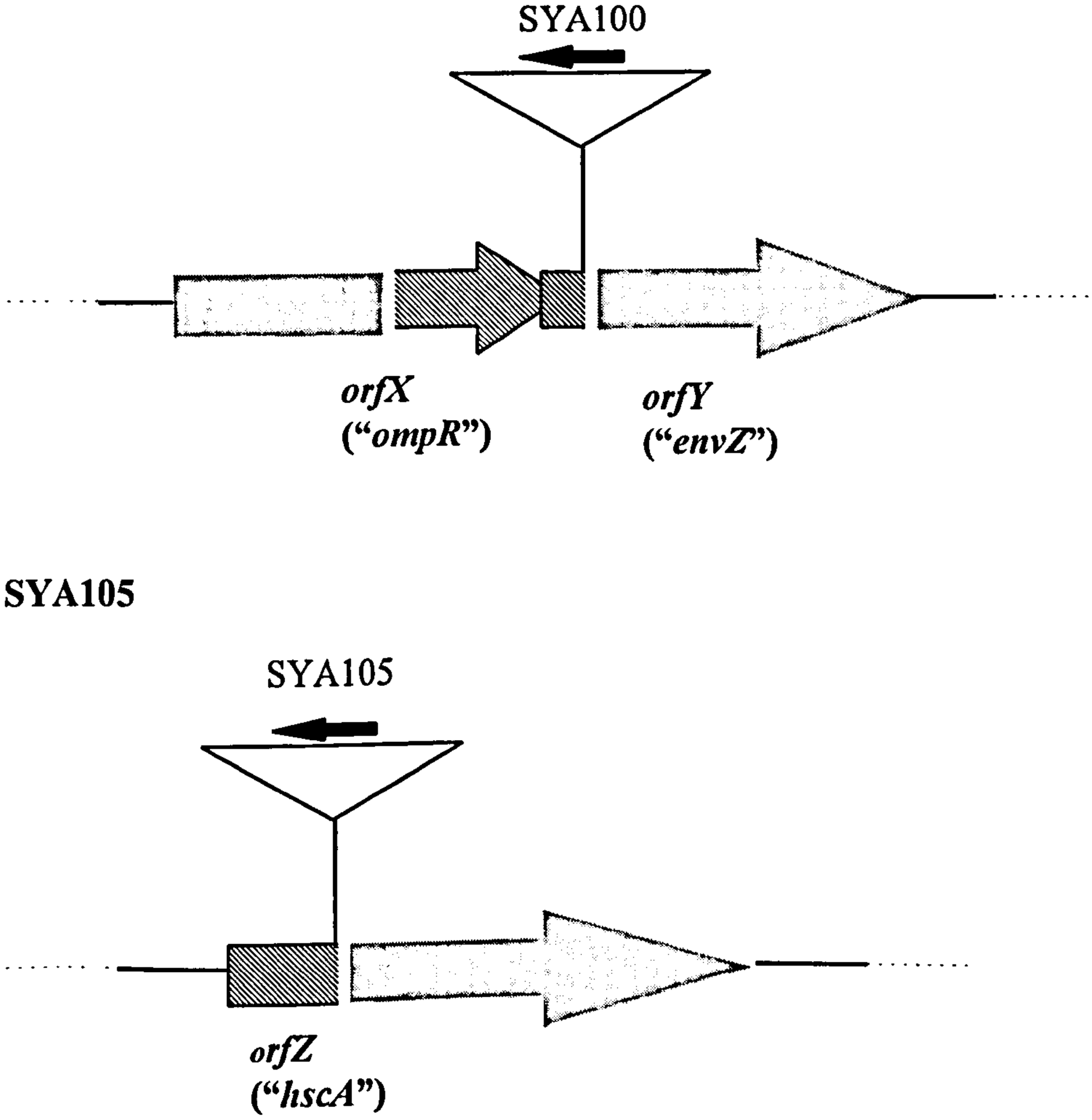


FIGURE 6.4 DEDUCED INSERTION POINTS OF Tn*phoA*'-2 IN Pig⁻ STRAINS SYA100 AND SYA105

Large arrows (not to scale), represent probable ORFs. Hatched shading represents regions in translated nucleotide sequences from the transposon clone of each strain which are homologous to database proteins. Coding regions designations (e.g. *orfX*) are shown below the respective regions, together with corresponding database homologues. Solid black arrows above Tn*phoA*'-2 insertions indicate the transcriptional orientation of *lacZ*, and hence the transposon, with respect to the deduced orientation of the coding regions. See text in section 6.3.2 for details.

6.4.3 DISCUSSION OF TRANSPOSON CLONE SEQUENCING RESULTS

Alignments of the sequences mapping to the *pig* cluster, together with database search information revealed the orientation of Tn*phoA*'-2, with respect to the transcriptional orientation of disrupted chromosomal regions in each strain. This suggested that both SYA101, SYA111 and SYA113 are gene fusion strains (see Fig. 6.3a). Both SYA111 and SYA113 were isolated as fusion strains from Tn*phoA*'-2 mutagenesis of SLAC1 (section 5.7.3). Sequence data therefore confirmed these observations. When SYA101 was transduced into the Lac^- background, it was also found to be Lac^+ (discussed in section 6.6.2), which confirmed the results presented in Table 6.1 for this strain. Sequence from construct pACYC177/102 was found to be complementary to the coding strand of *orf3* and database homologues of the translated amino acid sequence were present in the -3 reading frame, demonstrating that this is a non-fusion strain. This was later confirmed by transduction into the Lac^- background (section 6.6.2). Homology to database proteins for constructs pACYC177/100 from strain SYA100and pACYC177/105 from strain SYA105 (see Table 6.1) was detected in reverse reading frames, suggesting they are non-fusion strains. This was confirmed by generalised transduction into the Lac^- background (discussed in section 6.6.2). As shown in Fig 6.3a and 6.3b, transposon insertions disrupting transcription of *orf1* in the *pig* cluster map to different regions. By alignment it was deduced that in SYA113, Tn*phoA*'-2 is in fact inserted 5' to the predicted translational start site of *orf1*.

Figure 6.4a shows the approximate insertion point of Tn*phoA*'-2 in SYA100. It can be seen that the T*npho*A'-2 was localised to region *orfY*, which putatively encodes a protein with strong N-terminal similarity to EnvZ. In characterised host systems, *ompR* and *envZ* are translationally coupled and comprise the *ompB* locus (Lilljeström *et al.*, 1988). The positions of *orfX* and *orfY* in which homologies were detected in the non-coding strand suggest that they might also be arranged in overlapping fashion; this however might also be due to frameshifting caused by sequence error, which cannot be excluded as a possibility without further sequencing work.

In *E. coli* and *Sal. typhimurium*, *ompR* and *envZ* encode a two component sensor-regulator system, where EnvZ is a histidine sensor kinase and OpmR is a response regulator (Wurtzel *et al.*, 1982; Comeau *et al.*, 1985; Lilljetsröm *et al.*, 1988). Both proteins are required for the proper expression of the outer membrane porin proteins OmpC and OmpF, which function as passive diffusion pores for small hydrophilic molecules (Wurtzel *et al.*, 1982). Genes for OmpF and OmpC are genetically unlinked to the *ompB* locus, and expression of these proteins is determined indirectly by the osmolarity of the growth medium (Lugtenberg *et al.*, 1976; Van Alphen *et al.*, 1976), via EnvZ and OmpR (Wurtzel *et al.*, 1982). The EnvZ/OmpR system is part of a family of proteins of diverse origins which function in pairs and through which environmental signals act to modulate gene expression in prokaryotes (Lilljetsröm *et al.*, 1988). OmpR modulates target gene expression by direct interaction with extracistronic regulatory regions of DNA; a C-terminal domain is responsible for DNA binding and contains a helix-turn-helix motif (Kondo *et al.*, 1997; Martinez-Hackert and Stock, 1997). OmpR has N-terminal homology to other prokaryotic response regulators (Lilljeström *et al.*, 1988);

EnvZ, the sensor component of the system, is a two-pass inner membrane protein which belongs to a family of bacterial histidine kinases (Lilljeström *et al.*, 1988). The C-terminal region, located in the cytoplasm, contains the histidine residue which phosphorylates OmpR by direct interaction (Roberts *et al.*, 1994; Egger and Inouye, 1997; Hsing and Silhavy, 1997); this region has strong similarity to other bacterial sensor proteins, for example NtrB of *Bradyrhizobium* spp. and *Klebsiella pnuemoniae* which is involved in regulation nitrogen fixation, CpxA of *E. coli* which is involved in regulation of diverse cellular processes and VirA, involved in regulation of pathogenicity in *Agrobacterium tumefaciens* (Nixon *et al.*, 1986). Studies have suggested that the transmembrane regions of EnvZ, which are similar to those of other chemotactic signal transducers in bacteria (Lilljeström *et al.*, 1988), are essential for signal transduction (Tokishita *et al.*, 1992).

Tn*phoA*'-2 insertion in strain SYA105 was localised to a likely coding region designated *orfZ*, which is highly similar to the N-terminal region of *hscA* of

E. coli and *Haemophilus influenzae*. In *E. coli*, *hscA* encodes a 66 KDa heat shock cognate protein which bears resemblance to DnaK in the same host (Kawula and Lelivelt, 1994; Seaton and Vickery, 1994; Vickery *et al.*, 1997). DnaK, in turn, is homologous to eukaryotic 70 KDa heat shock protein (hsp70, Craig *et al.*, 1993).

Heat shock proteins are molecular chaperones which assist in the folding and assembly of proteins (reviewed by Craig *et al.*, 1993); DnaK is involved in the regulation of the heat shock response which occurs when cells are shifted from low to high temperatures (Craig and Gross, 1991). This system has been extensively studied in *E. coli*, and it is believed that DnaK and accessory proteins DnaJ and GrpE are involved in negative regulation of the heat shock response (Straus *et al.*, 1990).

Until the discovery of HscA, which is also termed Hcs66, DnaK was the only known prokaryotic hsp70 homologue. In contrast to DnaK, it has been found that in *E. coli*, Hsc66 is actually involved in an analogous stress-response mechanism which is induced when cells growing at higher temperatures are shifted to low temperatures, e.g. 10°C (Lelivelt and Kawula, 1995). Studies have shown that in *E.* coli, *hscA* forms part of an operon: a *dnaJ* homologue, *hscB*, has been identified immediately upstream of *hscA* (Kawlua and Lelivelt, 1994) and a ferredoxin-encoding *fdx* gene is found immediately downstream (Seaton and Vickery, 1994). It is hypothesised that Hsc66 might interact in some way with Fdx, possibly assisting in protein folding or assembly of the iron-sulphur cluster (Seaton and Vickery, 1994). HscB, a 20KDa protein which is also termed Hsc20, is a molecular chaperone containing a DnaJ-like domain which interacts with HcsA (Hsc66) and stimulates its ATPase activity (Vickery *et al.*, 1997). It has been suggested that sequence differences between Hsc66 and Hsc20 compared with other members of the chaperone superfamily might result in different peptide binding specificity and also that the proteins might be subject to different regulatory mechanisms (Vickery *et al.*, 1997). Hsc66 and Hsc20 have a high level of constitutive expression, suggesting that they play important cellular roles under non-stress conditions (Vickery *et al.*, 1997). Interestingly, no homology to *hscB* was detected in the nucleotide sequence of pACYC177/105, even though Tn*phoA*'-2 is inserted close to the N-terminus of *orfZ* (Fig. 6.3).

6.4.4.1 RE-ANALYSIS OF TRANSPOSON PROBING RESULTS

Restriction mapping and sequencing of the *pig* cluster had revealed the locations of *Eco*RI sites in this region of the *S. marcescens* chromosome (shown previously in Fig 3.1). Cloning and sequencing of transposon insertions had shown Tn*phoA*'-2 insertions in SYA101, SYA102, SYA111 and SYA113 map to the cluster, and their insertions points, within a narrow range, had been deduced (see Table 6.1). Using this information, results of transposon probing (section 6.3.2 and Fig. 6.1) were re-analysed to see if band sizes in these mutants are of expected sizes.

Tn*phoA*'-2 has a unique *Eco*RI site at 5640nt (see Fig. 5.1, Chapter 5). *Eco*RI digestion therefore cuts this transposon into two fragments. Predicted fragment sizes for every Tn*phoA*'-2 insertion which maps to the *pig* cluster were calculated using this information, and data compared with the band sizes in relevant lanes in Fig. 6.1, which were estimated from standard curves, by plotting log distance migrated (mm) against size (bp) (data not shown). Results of calculations are summarised in Table 6.2.

TABLE 6.2 PREDICTED AND DETECTED BAND SIZES OF Pig⁻ Tn*phoA*'-2 MUTANTS DIGESTED WITH *Eco*RI AND PROBED WITH TRANSPOSON-SPECIFIC DNA

LANE (Fig 6.1)	STRAIN	PREDICTED *Eco*RI BANDS[a] (APPROX. Kb)	*Eco*RI BANDS DETECTED BY PROBING[b] (Kb)
1	SYA100	n/a[c]	n/a
2	SYA101	13.1 + 10	>12.2 + 10.5
3	SYA102	14 + 11	>12.2 + 10.5
4	SYA105	n/a	n/a
5	SYA113	14 + 11	>12.2 + 10.9
6	SYA111	18.6 + 7	>12.2 + 7.8

[a]: Sizes of bands were calculated on the assumption that Tn*phoA*'-2 insertions in these mutants are at the lower limit of the insertion range predicted by sequence alignments. Ranges in which Tn*phoA*'-2 is located in each mutant are given in Table 6.1 and Fig 6.3b.
[b]: Bands of sizes >12.2 Kb could not be accurately estimated because the largest band in the 1Kb Ladder marker is equivalent to 12.216 Kb (see section 23.4.6).
[c]: n/a (not applicable) indicates that it is not possible to predict band sizes as these insertions lie outside the *pig* cluster in other loci.

6.4.4.2 CONCLUSIONS FROM TRANSPOSON PROBING RE-ANALYSIS

Table 6.2 shows that sizes of bands detected by probing mutants mapping to the *pig* cluster with transposon-specific DNA, are broadly equivalent to sizes which are predicted from calculating approximate insertion points in the *pig* cluster using nucleotide sequence data. This provided further evidence that transposon insertions are within the regions predicted by sequencing of transposon clones. Physical confirmation of this was then obtained by probing Pig$^-$ strain chromosomal DNAs with a different probe (section 6.5)

6.5.1 PROBING FOR TRANSPOSON INSERTIONS IN THE *pig* CLUSTER OF Tn*phoA*'-2 MUTANTS

Transposon cloning experiments had shown that Pig$^-$ phenotypes of different strains were attributable to insertional inactivation of relevant genes by Tn*phoA*'-2. Since sequence data suggest that two of the six mutations resulting in a Pig$^-$ phenotype lie outside of the *pig* cluster, it was decided to investigate this further.

A probe was made from *S. marcescens* DNA isolated from the prodigiosin biosynthetic cosmid pNRT104. It was predicted that if a transposon insertion were present in the *pig* cluster of a Pig$^-$ mutant, then its banding profile would differ from wild-type (and from strains with mutations outside of the *pig* cluster) if probed with *pig* cluster DNA. A Tn*phoA*'-2 insertion in the *pig* cluster should be apparent by the absence of the affected band at the same relative position as wild type, and the appearance of other bands not common to wild-type.

The probe used was made by random hexamer priming of *Bam*HI digested fragments of *S. marcescens* DNA from cosmid pNRT104. DNA, using digoxygenin-11-dUTP (DIG) (Boehringer Mannheim, section 2.7.1). *Bam*HI fragments, separated from pSF6 vector DNA in pNRT104 by agarose gel electrophoresis and extracted by GENECLEAN II™ (section 2.3.5.7), were provided by A. Cox. A probe generated previously from these fragments was found to hybridise to the *Bam*HI fragments of *S. marcescens* DNA in pNRT104, confirming that these fragments are specific for *pig* cluster DNA (A. Cox, Pers. comm.)

Chromosomal DNAs of SYA100, SYA101, SYA102, SYA105, SYA111 and SYA113 were digested using *Bam*HI as described previously (section 2.3.4.1) and run out overnight on an agarose gel at 20V. Southern blot transfer of DNA to Hybond-N™ nylon membrane (Amersham) was carried out as described in section 2.7.2 and the filter was probed overnight at 65°C (section 2.7.3).

6.5.2 RESULTS OF SOUTHERN BLOTTING

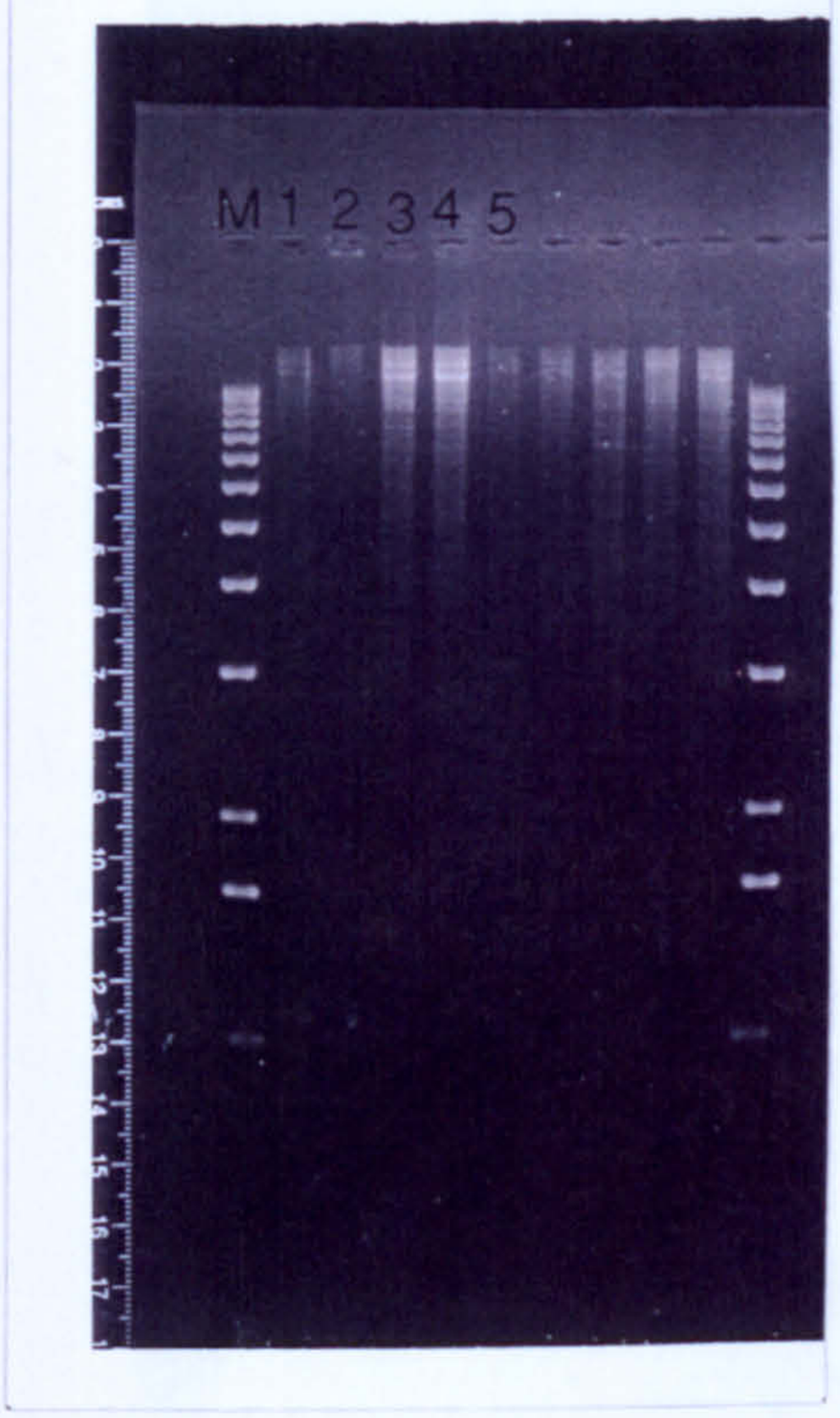
Results of Southern blotting are shown in Fig. 6.5. As predicted, Fig. 6.5 shows that banding profiles in certain lanes representing Pig⁻ mutants differ from wild-type (lanes 3, 4, 6 and 7). In lane 2 (SYA100), and lane 5 (SYA105), the banding profiles are the same as wild-type. This confirmed that Tn*phoA*'-2 insertions in SYA100 and SYA105 lie outside of the *pig* cluster. Bands of less than 1Kb were not detected in this blot (or when Southern blotting experiments were repeated, data not shown), although it is predicted that *Bam*HI digestion of *pig* cluster DNA should produce a fragment of 0.799 Kb. Absence of this band in all lanes is very likely due to poor signal detection from chromosomal DNA.

Absence of certain bands in lanes 3, 4, 6 and 7 is also accompanied by the appearance of additional bands in all the lanes except in lane 4. *Bam*HI digestion cuts Tn*phoA*'-2 into two fragments (see Fig. 6.2); insertion of additional (transposon) DNA changes the distances between formerly adjacent *Bam*HI sites in the chromosome of Tn*phoA*'-2 mutants, thereby altering banding patterns. In mutants SYA101, SYA102, SYA111 and SYA113, the absence of *Bam*HI fragment bands in comparison to wild-type, and the sizes of additional bands were noted. Band sizes were estimated by use of standard curves, plotting log distance migrated (mm) against size (bp) (data not shown). Results are presented in Table 6.3

FIGURE 6.5 RESULTS OF PROBING *Serratia marcescens*::Tn*phoA*'-2 Pig⁻ MUTANTS WITH A *pig* CLUSTER-SPECIFIC PROBE

Chromosomal digests and corresponding Southern blots are shown in Fig.6.5a and Fig. 6.5b respectively. *Bam*HI digested chromosomal DNAs were probed with a DIG-labelled pig cluster DNA from cosmid pNRT104. See text in sections 6.5.1, 6.5.2 and 6.5.3 for details and discussion of results

Figure 6.5a



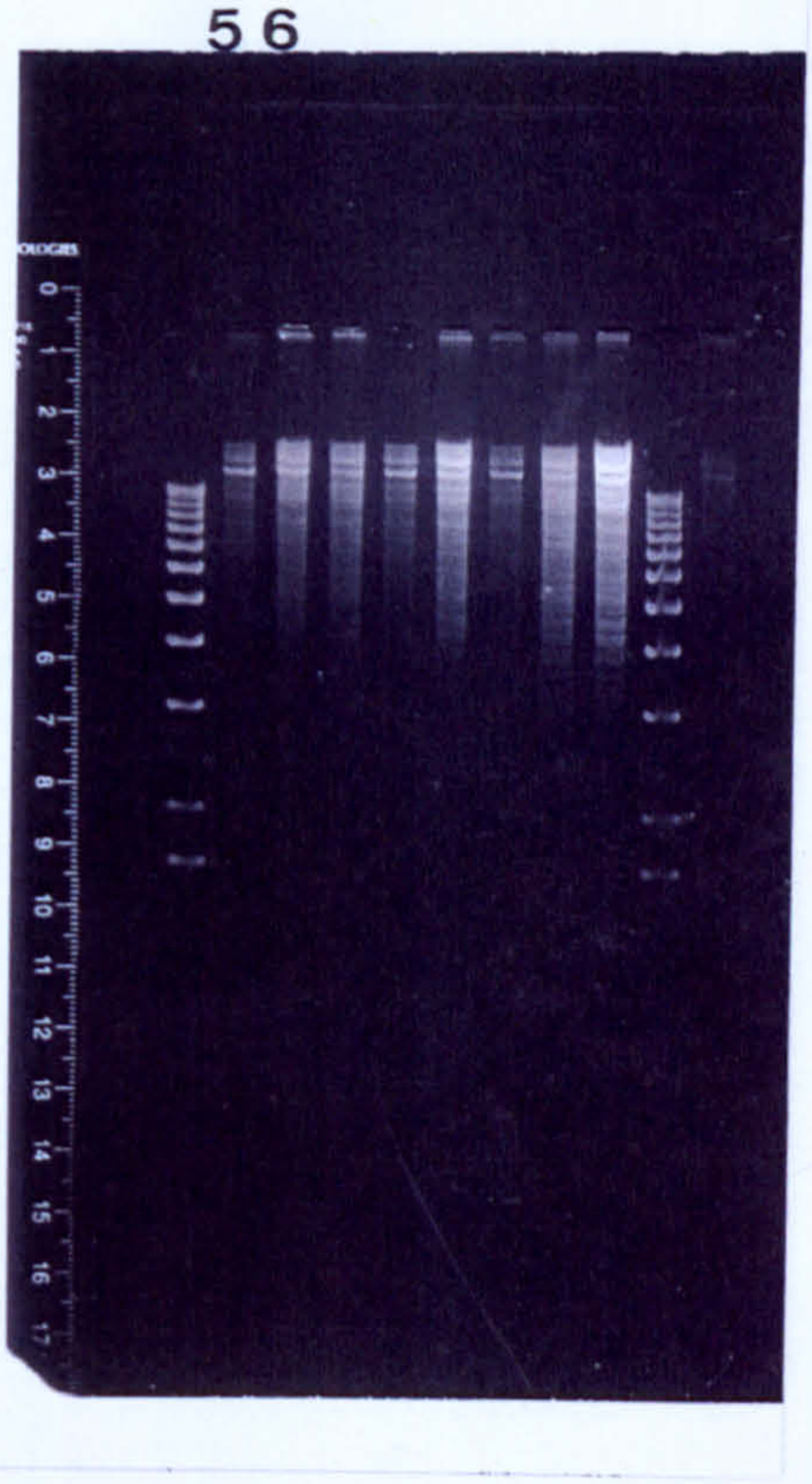


Figure 6.5a *Bam*HI digests of chromosomal DNAs
Key: M: 1Kb Ladder (Gibco BRL, see section 2.3.4.6); 1: SCAR1310124 wild-type; 2: SYA100; 3: SYA101; 4: SYA102; 5: SYA105; 6: SYA113; 7: SYA111

Figure 6.5b

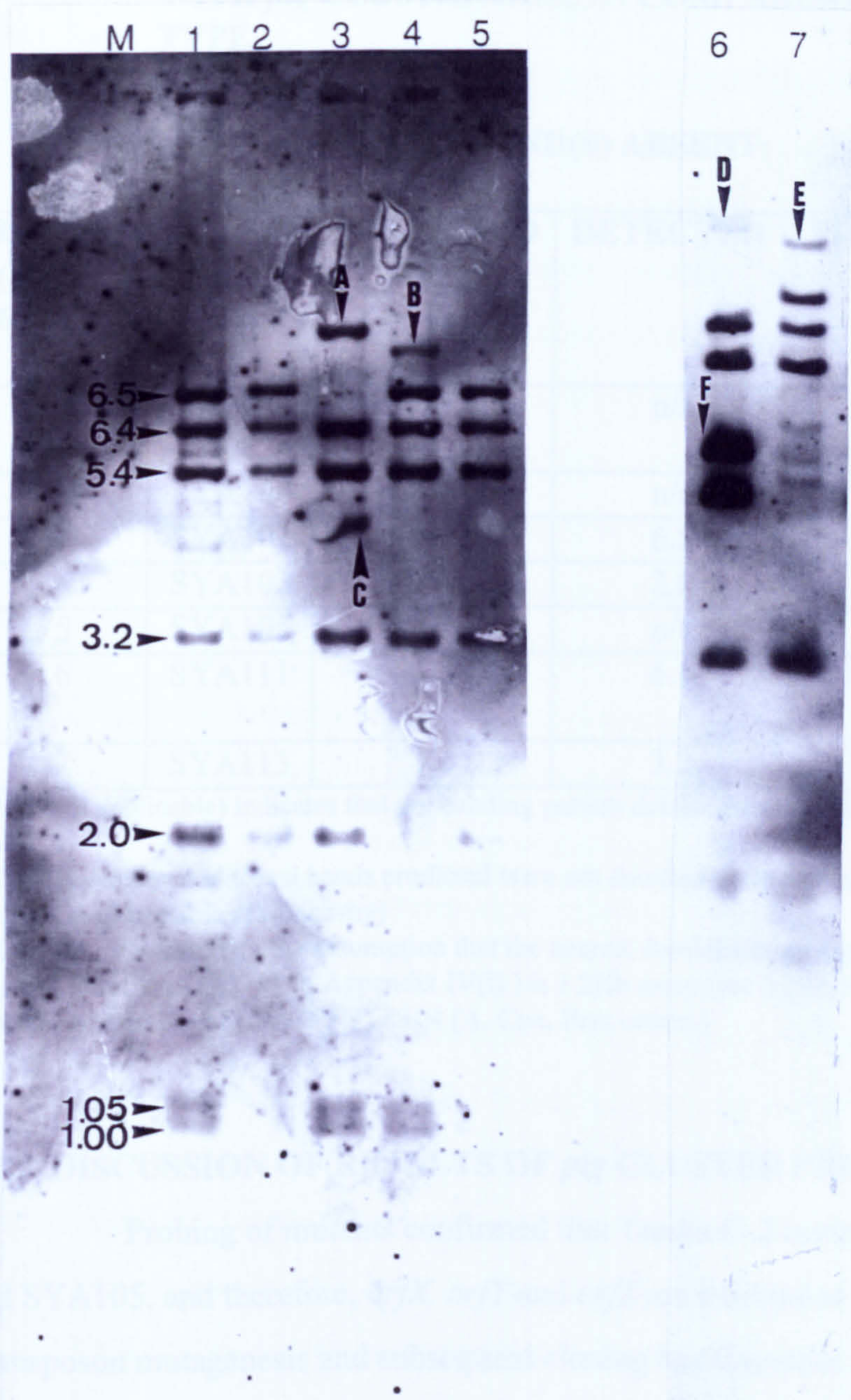


Figure 6.5b Southern blots of Tn*phoA*'-2 Pig⁻ mutants
Key: M: 1 Kb ladder (Gibco BRL, see section 2.3.4.6); 1: SCAR1310124 wild-type; 2: SYA100; 3: SYA101; 4: SYA102; 5: SYA105; 6: SYA113. 7: SYA111 Bands labelled A, B, C, D, E and F are additional bands resulting from Tn*phoA*'-2 chromosomal insertions, and they are discussed in section 6.5. The sizes (Kb) of wild-type *Bam*HI fragments of the *pig* cluster, as detected by the *pig* cluster DNA probe are labelled in lane 1.

TABLE 6.3 PREDICTED AND DETECTED BAND SHIFTS IN Pig⁻ STRAINS WITH Tn*phoA*'-2 INSERTIONS AFTER PROBING WITH *pig* CLUSTER DNA, IN COMPARISON TO WILD-TYPE

		*Bam*HI BAND(S) ABSENT: (Kb)		ADDITIONAL *Bam*HI BAND(S): (Kb)	
LANE (Fig. 6.5b)	**STRAIN**	**PREDICTED**	**DETECTED**	**PREDICTED**	**DETECTED (AND LABELLED IN Fig. 6.5)**
1	SCAR 1310124	n/a[a]	n/a	n/a	n/a
2	SYA100	n/a	n/a	n/a	n/a
3	SYA101	6.5	6.5	10.9 + 3.8	11(A)+ 4(C)
4	SYA102	2.0	2.0	9.2 + 0.9	9.2(B)+?[b]
5	SYA105	n/a	n/a	n/a	n/a
6	SYA111	6.5	6.5	11.6 + 3.1[c]	11.6(D) + 3.5(F)
7	SYA113	3.2	3.2	10.4 + 0.9	10.4 (E) + ?

a: n/a (not applicable) indicates that the banding pattern detected upon probing was that of wild-type.

b: ? indicates that additional bands predicted were not detected upon probing (see section 6.5.3 for discussion of possible reasons why)

c: This figure is based on the assumption that the nearest *Bam*HI site upstream of the nucleotide sequence start site (i.e., nt1 in Appendix IV(i)) is 1.2Kb away (see figure 3.1), as estimated by restriction mapping of cosmid pNTR104 (A. Cox, Pers comm.)

6.5.3 DISCUSSION OF RESULTS OF *pig* CLUSTER PROBING

Probing of mutants confirmed that Tn*phoA*'-2 insertions in SYA100 and SYA105, and therefore, *orfX, orfY* and *orfZ* are external to the *pig* cluster. Transposon mutagenesis and subsequent cloning has therefore identified two previously unknown loci which are required for prodigiosin biosynthesis.

Table 6.3 and Fig. 6.5 show that probing mutant strains with *pig* cluster DNA was largely successful in identifying regions of transposon insertions; the *Bam*HI fragment to which the insertion is localised in each case is, as predicted, absent in these mutants. The presence of additional bands which were detected in each mutant (labelled A-F in Fig. 6.5 and Table 6.3) are as would be predicted in each case, with the exception of band F, lane 6 in Fig. 6.5. Failure to detect 0.9 Kb bands in lanes 4 and 7 is most likely to be due to poor signal

detection, because the absence of a 0.8 Kb band in all lanes, which would represent the smallest *Bam*HI fragment of the *pig* cluster, is absent in all lanes, which suggests that small size fragments are missed by this method of detection. In lane 7, the predicted additional band of 3.1 Kb is not present, but an additional band of estimated size 3.5 Kb is present (labelled F, Fig.6.5b). With hindsight, it is possible that use of ^{32}P labelled probe, which is the conventional alternative to DIG labelling, might have circumvented the problems of poor chromosomal DNA yields causing reduced detection sensitivity which were experienced in blotting and hybridisation experiments, as radiolabelling provides a more sensitive means of detection. This approach might also have allowed detection of bands in digested chromosomal DNA preparations from strains SYA300 and SYA500. However, results of probing mutants with *pig* cluster DNA lend support to findings from transposon probing and cloning and sequencing experiments, particularly with regard to the locations of *orfX, orfY* and *orfZ*, which in mutants SYA100 and SYA105 do not map to the *pig* cluster. In these cases, it can be seen in Fig. 6.5 that banding patterns upon probing using *pig* cluster DNA are the same as wild-type (lanes 1 and 5).

Together, these experiments had established that transposon mutagenesis had been successful in inactivating and identifying genes relevant to prodigiosin biosynthesis. Strains SYA111 and SYA113 were known to be LacZ fusion strains, and this was confirmed by transposon cloning and sequencing. In order to confirm that SYA101 is also a fusion strain, and that SYA100, SYA102 and SYA105 are not, these mutants were transduced into the Lac$^-$ strain SLAC1.

6.6.1 SCREENING OF SCAR1310124::Tn*phoA*'-2 MUTANT STRAINS FOR LacZ FUSIONS

Generalised transduction by bacteriophages is an artefact of the packaging machinery which occurs when host DNA is encapsulated within a phage head during the lytic phase of the phage life cycle. Packaged DNA can then be transferred to a new host by whatever mechanism the phage uses. Occasionally, injected DNA recombines into the recipient host genome to form a stable recombinant (Masters, 1985). If the substituted allele possesses

a different phenotype from that of the inserted one, transductants can be screened for the novel phenotype. Localised mutagenesis, arising from homologous recombination, reduces the frequency of acquiring multiple mutations in the genome which is more likely to occur with other more general mutagenic strategies. Introduction of a donor allele that substitutes for the wild-type in the recipient cell creates a transductant clone that is isogenic with the recipient cell except for the exchanged allele. Of particular relevance in this study, a generalised transducing phage allows co-transduction of the transposon-induced gene mutation (or an adjacent gene of interest) with the transposon's selectable marker. In this study, transducing phage experiments demonstrated for certain mutant strains, that transposon insertions were responsible for the Pig^{-} phenotype. Generalised transduction was also attempted for Pig^{++} strain SYA500 and Pig^{O} strain SYA300.

ΦOT8 had been demonstrated to be capable of efficient generalised transduction between ATCC39006-derived *S. marcescens* strains without any apparent requirement for the presence of the LamB receptor or divalent cation supplement to growth medium (S. Bentley, Pers. comm.) and therefore it was decided to use this phage in this study.

6.6.2 ΦOT8 TRANSDUCTION OF SCAR1310124::Tn*phoA*'-2 Pig MUTANTS

ΦOT8 was provided by S. Bentley. In order to transduce Pig^{-}, Pig^{++} and Pig^{O} mutant strains into SLAC1, lysates were made on each of these strains. All experiments were performed using NBA plates and soft agar. Stock lysate was plaque purified, as described in section 2.5.1, on each mutant strain. This served to demonstrate, by the appearance of discrete pin-prick plaques, that the phage was capable of infecting each of these strains as discrete plaques are indicative of lysis from within. Using the plaque-purified phage suspensions, high titre lysates were made on each mutant strain by propagation (section 2.5.1). Lysates were harvested and titrated on their respective originating strains to ensure that they contained at least 10^9 pfu ml^{-1}, which would ensure a multiplicity of

infection (m.o.i) of at least 1 for overnight cultures during the subsequent transduction experiment.

Separate overnight cultures of SLAC1, grown in NB were transduced using each lysate as described previously (section 2.5.3). After washing and expression, transductants were plated on NBA-Tc-X-Gal to select for Tn*phoA*'-2 recombinants and screen directly for LacZ fusions; non-transduced aliquots of SLAC1 were also plated onto NBA-Tc-X-Gal as negative controls. Plates were incubated at 30°C for three days, except SYA300 and SYA500 which were incubated for four days at 37°C to prevent obscuring of blue-white screening by pigment production. Results are shown in Table 6.5 .

TABLE 6.4 RESULTS OF ΦOT8-MEDIATED TRANSDUCTIONS OF SCAR1310124::Tn*phoA*'-2 Pig MUTANTS INTO THE Lac⁻ BACKGROUND

SCAR1310124 ::Tn*phoA*'-2 STRAIN, (Tn*phoA*'-2 INSERTION) AND Pig PHENOTYPE	TOTAL $N^{Ω}$ Tc^{R} COLONIES	$N^{Ω}$ Pig^{-}	$N^{Ω}$ Pig^{+}	% EFFICIENCY OF Tc^{R}::Pig PHENOTYPE COTRANS-DUCTION	Lac^{+} or Lac^{-}
SYA100, (*orfX*), Pig^{-}	1300	1294	6	99.6	+/-
SYA101, (*orf1*), Pig^{-}	600	589	11	98.2	++
SYA102, (*orf3*), Pig^{-}	370	356	14	96	+/-
SYA105, (*orfZ*), Pig^{-}	880	841	39	95.6	-
SYA300, (?), Pig^{O}	0	-	-	-	n/a
SYA500, (?), Pig^{++}	0	-	-	-	n/a

6.6.3 DISCUSSION OF ΦOT8 TRANSDUCTION EXPERIMENTS

Results in Table 6.3 show that transductions were successful in all of the strains with the exception of the orange phenotype mutant SYA300 and hyperpigmented mutant SYA500. The preparation of lysate and the transduction procedure were attempted several times, without success, on Pig^{o} and Pig^{++} strains. Since ΦOT8 was observed to produce plaques on these strains, which allowed production of viable lysates, this excludes the possibility that the strains are altered in cell surface properties in some way which prevented phage adsorption and therefore transduction. Because mutated phenotypes were being transferred between *S. marcescens* strains, the possibility of the mutation being lethal can be excluded. One possible explanation for the failure to obtain transductant colonies for SYA300 and SYA500 is that transposon insertions might be unstable in these regions of the chromosome; if a transposon is excised from these sites, then Tc^{R} colonies would not be obtained by Tc selection subsequent to transduction. Due to the problems encountered in chromosomal DNA preparations for SYA300 and SYA500 discussed in section 6.2), these strains were not probed for Tn*phoA*'-2 so these observations remain inconclusive. For the other strains, transduction efficiencies of near 100%, i.e. co-inheritance of Tc^{R} with the parental strain pigment phenotype (Table 6.4), indicate that single transposon insertions were present in these strains, in agreement with transposon-probe hybridisation data (section 6.2). Co-inheritance of Tc^{R} with the Pig- phenotype also demonstrates that transposon insertions are responsible for the Pig^{-} in these strains. As transposon probing (section 6.3.1) had already shown that single transposon insertions are presented in Pig^{-} strains, the main objective of the generalised transduction experiments was to screen the mutagenised strains for potential gene fusions. Table 6.1 shows that out of the four strains transduced, SYA101, appeared to be unequivocally Lac^{+}. SYA105 produced white colonies on NBA-X-Gal and it was therefore concluded that it is not a fusion strain, confirming results of transposon clone sequencing (section 6.4.2). Strains SYA100 and SYA102 appeared faintly blue (Table 6.4). Sequencing had revealed that transposon insertions in SYA100 and SYA102 are in the opposite orientation to *orfX*, *orfY* and *orf3*, and this is supported by transposon and *pig* cluster probing

studies, so it was concluded that these mutants cannot be true LacZ fusion strains. Appearance of pale blue colonies resulting from Tn*phoA*'-2 mutagenesis has been also reported by the workers who constructed the transposon; this phenomenon is apparently due to low basal level expression of LacZ in this and similar transcriptional fusion constructs. (Wilmes-Reisenberg and Wanner, 1992) Generalised transduction was therefore successful in constructing one gene fusion strain for expression studies.

6.7 GENERAL DISCUSSION AND SUMMARY

The six Pig$^-$ strains generated by transposon mutagenesis were characterised further by experiments described in this chapter. The studies showed that genes required for prodigiosin biosynthesis are present in the *pig* cluster and also two previously unknown loci, outside the *pig* cluster, were identified.

The precise locations of transposon insertions are not known in these mutants. This could be resolved by firstly, cloning transposon and flanking chromosomal DNA adjacent to IS*50*L. This would exploit the unique *Eco*RI site in Tn*phoA*'-2, in the same way as the *Bam*HI site in the transposon was exploited for cloning insertions in this study (see Fig. 6.2). These cloning experiments were not done during this study because it was thought that cloning of insertions on *Bam*HI fragments had provided sufficient data to localise some of the insertions to the *pig* cluster and also to confirm orientation of the transposon insertion in each case. However, pinpointing precise locations of Tn*phoA*'-2 insertions in these mutants would an obvious and logical next step. Secondly, localisation studies involve the design and use of reverse primers to sequence *towards* IS*50* elements , which would identify, to the exact nucleotide, the precise insertion point of each Tn*phoA*'-2 in each strain.

Sequencing data showed that Tn*phoA*'-2 had inactivated two loci mapping outside of the *pig* cluster. Putative coding regions (*orfX, orfY* and *orfZ*) have very strong similarity to known enteric bacterial proteins. Homologues of OrfX, OrfY and OrfZ, (OmpR, EnvZ and Hsc66 respectively), function as part of gene expression regulatory systems in their cognate hosts. OmpR and EnvZ

regulate gene expression in response to changing osmolarity. The high degree of similarity which *orfZ* shares with the N-terminal region of Hsc66 is intriguing. Only three Hsc66-type proteins are currently catalogued in the SwissProt database. As discussed previously, *hscB* expression is raised in cold shock, but it has also been suggested that it might perform housekeeping roles under non-stress conditions. It is possible that a Hsc66 homologue might be required for correct assembly of some of the prodigiosin biosynthetic proteins, and insertional activation of *orfZ* therefore results in misfolding of these proteins. As mentioned previously (section 1.3.6.3), prodigiosin biosynthesis does not occur at temperatures <12°C, so the identification of a putative "cold-shock" protein is also very interesting in this context. In *E. coli*, a co-chaperone (Hsc20) encoding gene, *hscB*, has been identified immediately upstream of *hscA*. Database searches with the Tn*phoA*'-2::*orfZ* nucleotide sequence failed to find any homologues for the region of sequence extending beyond *orfZ*. Therefore, on a cautionary note, identification of *orfX, orfY* and *orfZ* as *bona fide* homologues of *ompR*, *envZ* and *hcsA* cannot be done until more sequence data is obtained. Since these putative ORFs are not transcriptionally fused to *lacZ* in their respective host strains, they could not be used in physiological studies and therefore they were not investigated further in this study. However, there is much scope for future work with these mutants. Firstly, sequencing of DNA flanking IS*50*L in these mutants would be very informative in characterising the loci further. Further sequencing, by designing reverse primers and sequencing towards IS elements of Tn*phoA*'-2 would enable precise pinpointing of transposon insertions. Further outward sequencing, again by designing primers should be done before assigning possible roles to *orfX, orfY* and *orfZ* on the basis of similarity.

It would then be essential to identify wild-type *orfX, orfY* and *orfZ* in *S. marcescens*. This could begin by carrying out complementation studies using the *S. marcescens* cosmid library, which is available in this laboratory, and screening for restoration of pigment production in these mutants. The transposon-

based studies presented in this and the previous chapter which resulted in identification of *orfX, orfY* and *orfZ* have therefore identified exciting avenues for future investigation.

From these experiments, three fusion strains were generated which were used in LacZ expression studies (presented in Chapter7). In SYA102, Tn*pho*A'-2 was shown to be inserted in *orf3* in the *pig* cluster, but this is not a fusion strain and was therefore not studied further. If it were a fusion strain, it would have been interesting to see if there is any difference in the LacZ expression profile in comparison with SYA101 and SYA113, as this would have provided clues to the transcriptional organisation of these ORFs. However, it is noted that transposon mutagenesis causes gross disruption of coding regions, so if putative ORFs are translationally coupled in the *pig* cluster, it is quite possible that transposon insertions exert polar effects on transcription of downstream genes.

CHAPTER 7

REPORTER GENE STUDIES ON PRODIGIOSIN BIOSYNTHESIS

7.1 PREFACE

Strains SYA113, SYA1011 and SYA111 are gene fusion strains which map to the *pig* cluster. In the final part of the study, experiments were done to assess gene activity by LacZ expression studies. Due to time constraints, the work which was done was limited in nature, but nonetheless raised some interesting findings for possible future investigation. The factors which reportedly influence prodigiosin biosynthesis in *S. marcescens* were discussed in section 1.3.6. However, these data can be regarded as phenomenological, because the effect of varying environmental conditions on prodigiosin biosynthesis has not been studied at the genetic level. The construction of gene fusion strains was therefore an invaluable tool in this respect.

7.2 AIMS

The aim of this work was to score reporter gene activity under varying environmental conditions in the different strains, and to asses whether there is any differential expression of *lacZ* between the different strains.

7.3.1 THE INFLUENCE OF TEMPERATURE ON LacZ EXPRESSION IN *pig* CLUSTER GENE FUSION STRAINS

The reported influence of temperature was reviewed in section 1.3.6.3. It is well established that prodigiosin biosynthesis is arrested at temperatures >36°C. Firstly it was decided to follow LacZ expression through growth of the strains at a permissive temperature (30°C). This experiment would also show if there are any differences between gene expression between the fusion strains.

Culture flasks for strains SCAR1310124 (wild-type, Lac^+), SLAC1 (wild-type, Lac^-), SYA1011 (*orf1*), SYA113 (*orf1*) and SYA111 (*orf4*) were set up as described previously (section 2.10). Strains were cultured in 25 ml of LB in 250 ml flasks at 30°C with good aeration (200 rpm) and a starting OD_{600}= 0.05. 2x 1 ml samples of culture were removed from each flask at 2h intervals. One sample was used to measure OD_{600}, and the other sample was subjected to toluene treatment (section 2.12.1). The latter samples were then assessed for LacZ (β-galactosidase) activity as described in section 2.12.2. The experiment was

galactosidase) activity as described in section 2.12.2. The experiment was repeated under the same conditions and results were found to be the same. Results are presented in Fig.7.1.

7.3.2 ANALYSIS OF LacZ ASSAY RESULTS

Fig. 7.1 shows that growth rates for all strains are the same. From these data, doubling time in log phase was calculated to be 2 hours for the strains. Results show very obvious differences in LacZ activity between certain strains. Firstly, it can be seen that, as stated previously, there is negligible LacZ activity in the Lac$^-$ wild type SLAC1. In contrast to this, the fusion strains show marked LacZ activity. It can be seen that strains SYA1011 and SYA113 have the same LacZ profile through the growth curve. This is because they are both inserted in or adjacent to the *orf1* region of the *pig* cluster. Tn*phoA*'-2 is inserted approximately 15nt upstream of the proposed translational start site of *orf1* in SYA113, whereas it is inserted approximately 615nt into *orf1* but this has not affected LacZ activity, and this is because they are transcriptional fusions, with *lacZ* having its own RBS. The fact that activity is the same in both strains also provides other clues to the nature of Orf1: Orf1 is likely to be a cytoplasmic protein; if it is a transmembrane protein, Tn*phoA*'-2 insertions in SYA1011 and SYA113 are likely to be in cytoplasmic domains of the protein. This inference is drawn from the fact that β-galactosidase is not an export competent protein: if a gene fusion results in LacZ being directed to the cell envelope, enzymatic activity is greatly reduced; if a signal sequence precedes LacZ in the hybrid protein (which obviously cannot be the case in SYA113), then lethal jamming of the cell's export machinery can occur, i.e. in such fusions, it results in overproduction lethality (Slauch and Silhavy, 1991).

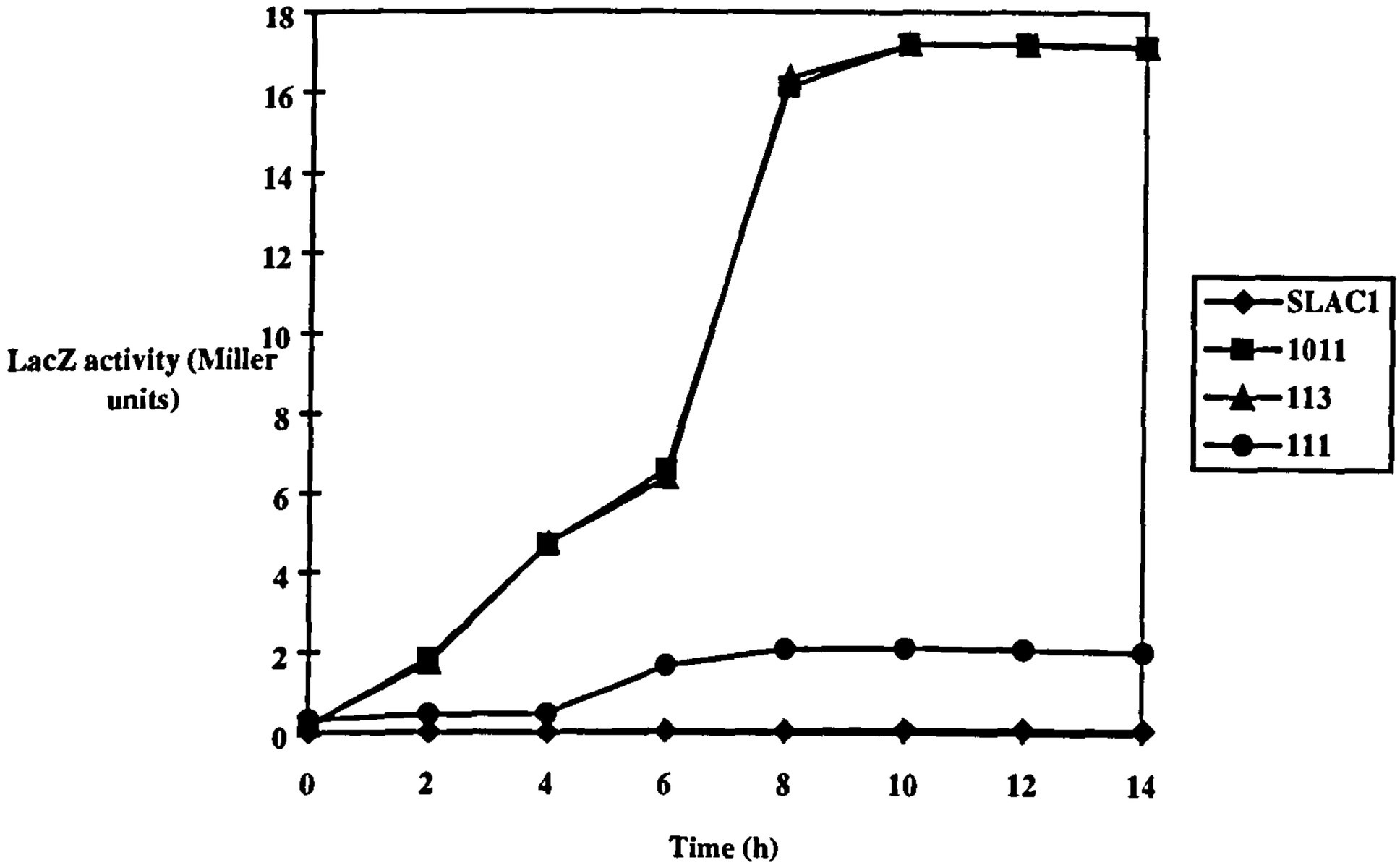
Comparison of the LacZ profiles of SYA1011 and SYA113 to that of SYA111 shows dramatic differences in the profile and amount of LacZ activity. These results suggest that *orf1* is expressed more highly than *orf4* under these experimental conditions. This could be because *orf4* is expression is subject to differential regulation at the DNA level to *orf1*, which therefore suggests that the 163nt gap between *orf3* and *orf4* may be important in downstream gene

FIGURE 7.1 β-GALACTOSIDASE ACTIVITY OF Pig⁻ Tn*phoA*'-2 TRANSCRIPTIONAL FUSION STRAINS AT 30°C

β-galactosidase activity (**A**) was measured through the growth curve by the method of Miller *et al.*, (1972, section 2.12.2). The growth curve (**B**) was drawn using OD_{600} values of samples taken at 2h intervals on a Helios Alpha spectrophotometer (Unicam). SLAC1 is the Lac⁻ wild-type control. All other strains are Tn*phoA*'-2 Pig⁻ mutants (SYA1011 etc.), which are LacZ fusion strains, generated as described in previous chapters. See text in section 7.3.2 for discussion of results.

Figure 7.1

A


18
16
14
12
10
8
6
4
2
0
LacZ activity (Miller units)
0
2
4
6
8
10
12
14
Time (h)
SLAC1
1011
113
111

B

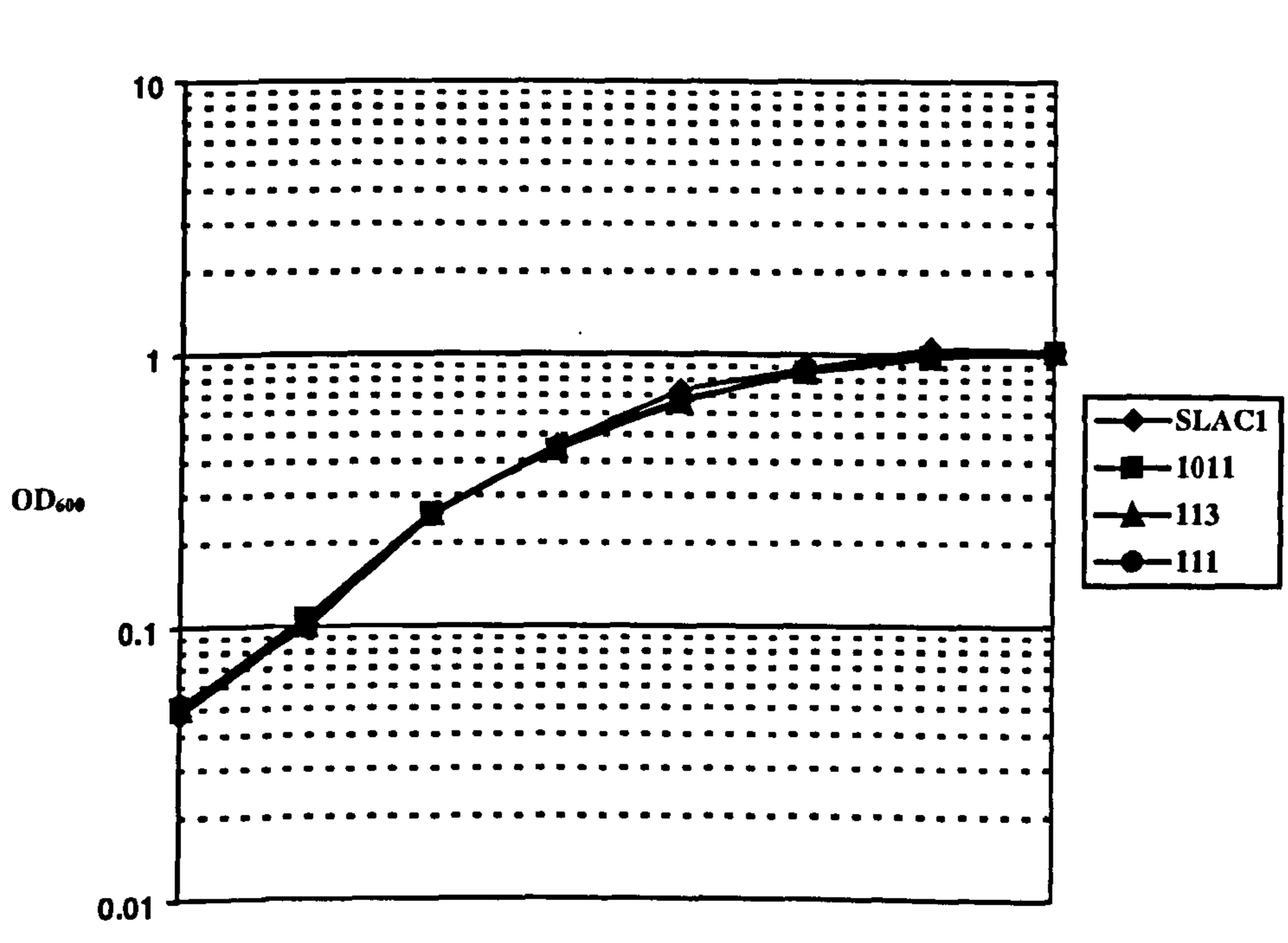
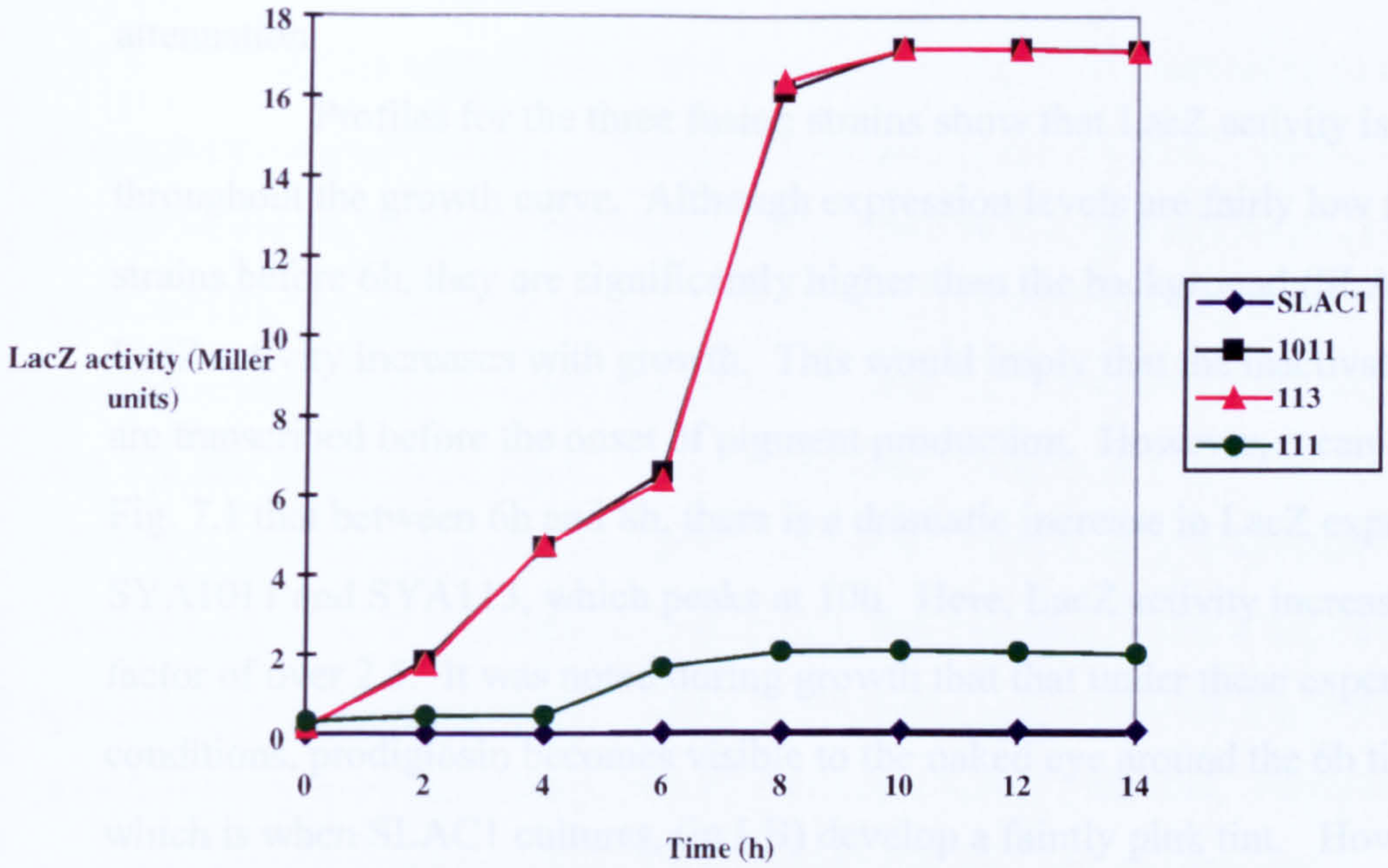
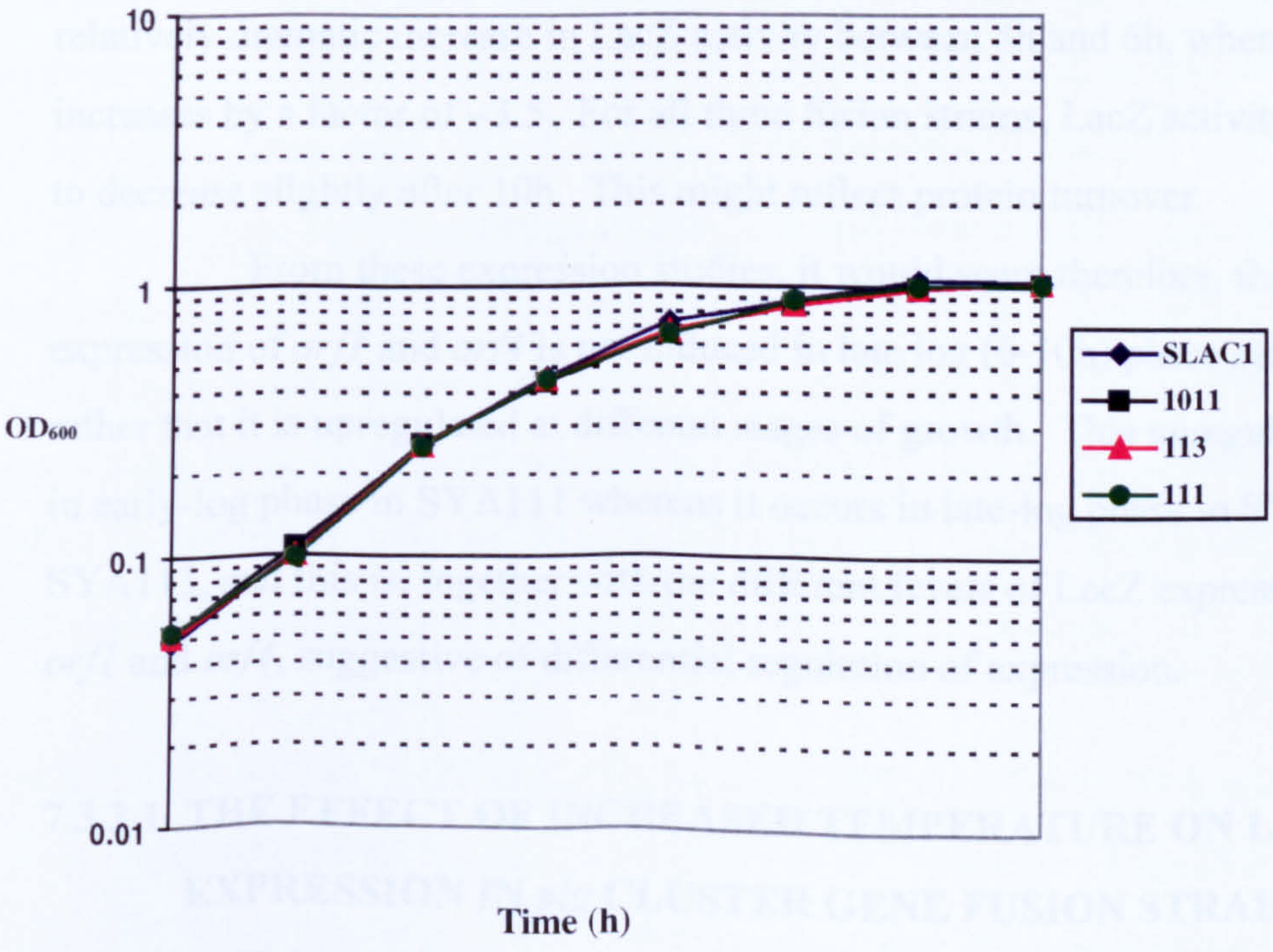

10
1
0.1
0.01
OD600
Time (h)
SLAC1
1011
113
111

Figure 7.1

A


18
16
14
12
10
8
6
4
2
0
LacZ activity (Miller units)
0
2
4
6
8
10
12
14
Time (h)
SLAC1
1011
113
111

B


10
1
0.1
0.01
OD600
SLAC1
1011
113
111
Time (h)

expression. If *orf1* and *orf4* are both dependent upon the same promoter, the differences in LacZ activity might be due to other regulatory factors, for example attenuation.

Profiles for the three fusion strains show that LacZ activity is present throughout the growth curve. Although expression levels are fairly low for the strains before 6h, they are significantly higher than the background (SLAC1), and LacZ activity increases with growth. This would imply that the inactivated genes are transcribed before the onset of pigment production. However, it can be seen in Fig. 7.1 that between 6h and 8h, there is a dramatic increase in LacZ expression in SYA1011 and SYA113, which peaks at 10h. Here, LacZ activity increases by a factor of over 2.5. It was noted during growth that that under these experimental conditions, prodigiosin becomes visible to the naked eye around the 6h time point, which is when SLAC1 cultures, (in LB) develop a faintly pink tint. However, it was also noted that between the 0h and 2h time points, LacZ expression in SYA113 and SYA1011 increases by a factor of >10. In SYA111, there is a relatively dramatic increase in LacZ activity between 4h and 6h, when, activity increases by a factor of ~3.5. For all three fusion strains, LacZ activity was seen to decrease slightly after 10h. This might reflect protein turnover.

From these expression studies, it would seem therefore, that that expression of *orf1* and *orf4* is not induced in late log (6-10h) phase *per se*, but rather that it is upregulated at different stages of growth. This upregulation occurs in early-log phase in SYA111 whereas it occurs in late-log phase in SYA1011 and SYA113, and this is, together with the different levels of LacZ expressed from *orf1* and *orf4*, suggestive of differential regulation of expression.

7.3.3.1 THE EFFECT OF INCREASED TEMPERATURE ON LacZ EXPRESSION IN *pig* CLUSTER GENE FUSION STRAINS

Since prodigiosin biosynthesis is abolished at temperatures >36°C, it was decided to see if this results in down-regulation or abolition of LacZ expression in the fusion strains. To assess if LacZ activity is simply "on" or "off" at a restrictive temperature, liquid cultures of fusion strains and SLAC1 were

grown in 10 ml of LB, overnight at 30°C as described previously (section 2.1.2). 50µl of each overnight culture was then added to 2 x 5 ml of LB (in 25 ml universal tubes). for each strain, one tube was placed again at 30°C as before, and one was placed at 37°C (as described previously, section 2.1.2). All tubes were incubated for 48h. Following this, 5µl of each culture grown at 37°C was spotted onto LBA-X-Gal and incubated for 4 days at 37°C. As positive control, cultures grown at 30°C strains were assayed in the same way at 30°C, again for four days.

7.3.3.2 RESULTS OF LacZ ASSAYS OF STRAINS CULTURED AT 37°C

Using this crude indicator assay, it was found that all there was no apparent difference in the intensity of colour of the strains between 30°C and 37°C (data not shown). This therefore indicates that that *orf1* and *orf4* expression is not abolished by temperature.

7.3.3.3 ANALYSIS OF RESULTS

It is hypothesised that one of the enzymes responsible for condensation of MAP and MBC to form prodigiosin is temperature sensitive (section 1.3.6.3). These results suggest that transcription of *orf1* and *orf4* is not sensitive to temperature. This experiment was useful in providing preliminary data, but gene expression at the restrictive temperature should, in the future, be followed through the growth curve to assess if the expression profile is altered, even though it is not abolished, as this experiment showed.

7.4 THE EFFECT OF INACTIVATING THE POSITIVE REGULATOR Rap

7.4.1 PREFACE

The generalised transducing phage ΦOT8, was used in this experiment to create a double mutant strain. The phage was propagated on a Rap$^-$ strain which had been constructed previously by N. Thomson (1996). The resultant lysate was used to transduce the *orf1*::*lacZ* fusion strain SYA1011, and the resulting strain was then assessed for LacZ activity. In the time remaining in the final stages of

the study, this experiment was performed successfully on strain SYA1011. There was insufficient time to perform the experiment on the other gene fusion strains generated by Tn*phoA*'-2 mutagenesis.

7.4.2 CONSTRUCTION OF STRAIN NTSA1

A ΦOT8 lysate was made on strain NTM2 as described previously (2.5.3), after the phage had been plaque-purified on this strain (section 2.5.3). NTM2 was constructed by marker exchange mutagenesis, and carries a *rap* gene that is insertionally inactivated by a Kn^R casette. In NTM2, *rap* is truncated to to 42 amino acids; wild-type *rap* is 145 amino acids in length (Thomson, 1996). After performing a phage titration assay using NTM2 (section 2.5.1) to ensure that $>10^9$ pfu ml^{-1} are present in the lysate, it was used to transduce strain SYA1011, as described previously (section 2.5.3 and 2.5.2). Following expression, transductants were selected on NBA-Kn-Tc-X-Gal plates at 30°C. A non-transduced aliquot of SYA1011 was also plated as above as negative control. Use of X-Gal in the growth medium would provide a direct indication of LacZ expression; Kn selected for *rap*::Kn recombinants, while Tc selected for Tn*phoA*'-2.

7.4.4 RESULTS OF TRANSDUCTION OF SYA1011

The total number of transductants obtained in this experiment was 287. Kn^R/Tc^R colonies which appeared on selective agar were presumed to be *rap/orf1* double mutants. All the colonies appeared white on the selective agar (data not shown). The strain was designated as NTSA1. Negative control plates failed to yield any colonies, implying that double-mutant transductants had been generated by homologous recombination. The strain derived from this experiment was designated NTSA1.

7.4.4 INTERPRETATION OF LacZ PHENOTYPE SCORING

The appearance of white, Kn^R colonies on X-Gal agar suggests that LacZ activity was abolished by insertional inactivation of *rap*. This strain is constructed from SYA1011, which is presumed to carry Tn*phoA*'-2, and therefore

lacZ (see section 6.6.2), within *orf1* of the *pig* cluster. This results therefore suggests that activity of *orf1* is dependent upon positive regulation by the product of *rap*. As *orf2* and *orf3* are predicted to be translationally coupled to *orf1*, it would be predicted that Rap is also essential for their expression. NTSA1 was also tested for carbapenem activity using *E. coli* ESS indicator plates, as described previously (section 2.9), using SLAC1 and SCAR1310124 as positive controls and NTM2 as a negative control. Carbapenem production was not evident by this assay in NTSA1 or NTM2.

Due to time constraints, no further work was possible on this strain. In order to physically confirm that *rap* has been inactivated in NTSA1, PCR should be performed using *rap* primers Cyto1 and Cyto2, as was done previously in other Pig$^-$ mutant strains (see section 5.5.3.1). In NTSA, it would be predicted that the PCR-generated product would be ~1740bp, which corresponds to this size of *rap* containing a 1266bp Kn^R cassette (Thomson, 1996), and would be the same as that obtained by performing PCR on NTM2 DNA using these primers. In order to confirm that LacZ activity is abolished in NTSA1, spectrophotometric assays for β-galactosidase are also essential in order to quantitate the result. Performing this experiment in other Tn*phoA*'-2-generated gene fusion strains (SYA111 and SYA113), would also be interesting as it would provide more information on the regulation of the *pig* cluster. It would be predicted that LacZ activity would also be abolished in strain SYA113, in which a Tn*phoA*'-2 insertion is present immediately upstream of *orf1* of the *pig* cluster. If LacZ remains active after performing this experiment in strain SYA111, which carries a Tn*phoA*'-2 insertion in *orf4*, this would provide evidence of differential regulation of this gene (and possibly downstream genes translationally coupled to *orf4*), which would possibly implicate the small intergenic region between *orf3* and *orf4* in regulation of pigment biosynthesis in this species.

CHAPTER 8

CONCLUDING REMARKS AND FUTURE WORK

8 CONCLUDING REMARKS AND FUTURE WORK

The overall aim of this study was to investigate the biosynthesis of prodigiosin at the genetic level. At the outset of this study, virtually no published information existed about the nature of the genes involved in this process, or the products of these genes. It was estimated that the *S. marcescens pig* cluster is between 15-25Kb in size. The smallest functional subclone of pNRT104 which was generated was plasmid pM245, which is 25.056 Kb in size. Approximately half of the *Serratia* DNA carried in this construct was sequenced in this study, and analysed further and mutated to see what effects this has on the pigment phenotype and the physiology of the host system. The remaining *pig* DNA was sequenced in a parallel study in this laboratory (A. Cox, unpublished). Conclusions are set out below, together with suggestions for future work.

Sequencing of the *pig* cluster revealed that putative ORFs are arranged undirectionally, with the exception of the terminal putative ORF, *orf16*. Consequently, *pig* DNA sequenced in this study was designated the 5' end of the cluster. Sequence analysis revealed 6 putative ORFs in this region of *pig* DNA. The first of these is a partial ORF, of which only the terminal 3' end is present and encodes a NADH peroxidase homologue. The first designated ORF of the *pig* cluster encodes an acyl CoA dehydrogenase homologue. *orf2* encodes a protein which is homologuous to a *Str. coelicolor* protein of unknown function which maps to the *red* cluster in that species. *orf3* encodes a PEP synthase homologue. *orf4* is a hypothetical protein of unknown function which does not have, at present, any database homologues. *orf5* encodes a homologue of ornithine aminotransferase. The putative gene arrangement of these ORFs is one of apparent translational coupling. *orf1*, *orf2*, and *orf3*, and *orf4* and *orf5* are arranged in two units, separated by a small intergenic region. Screening for promoter elements identified two regions of similarity to sigma-70 type promoters by sequence conservation. Screening for motifs of alternative sigma-factor dependent promoters failed to reveal any obvious areas of resemblance. On basis of this data, it is not possible to ascertain if this putative gene arrangement

represents two or one polycistrons. Intergenic regions were found to be abundant in sequence with the potential to form stem-loop structures and many direct repeats were also identified. This suggests that these regions do play a role in regulating *pig* cluster gene expression. Attempts to express putative proteins in the T7 expression system was partly successful. The failure to obtain bands of predicted sized in all cases may be due to effects of translational coupling, poor expression in *E. coli* of *pig* genes, or it may be that the putative coding regions are in reality, arranged differently to the predicted arrangement. For example, it is possible the *orf5* is actually two proteins but appears to be one larger polypeptide because of an undetected frameshift error in the sequence data. As stated previously, this work therefore shows that assignment of coding regions must be treated with caution when done on the basis of sequence data alone. Further work is necessary to physically confirm putative start and stop sites in coding regions. For unequivocal identification of start sites, further work should involve primer extension studies, S1 nuclease mapping and N-terminal sequencing of isolated proteins. Information derived from primer extension analyses would be extremely useful in future attempts to subclone the cluster and express the proteins. As an alternative to T7-expression, proteins could be expressed *in vitro*, for example using the minincell system. This would overcome any possibility of poor expression due to the *E. coli* heterologous host system.

In order to confirm that each putative ORF identified by sequencing of the *pig* cluster is essential to prodigiosin biosynthesis, it would be necessary to mutate it and assess the impact this has on pigment phenotype. As sequence data were not available to employ a directed approach, random (transposon) mutagenesis was used to isolate genes required for pigment production in this study. This resulted in generation of non-pigmented mutants in which the causative lesions mapped to *orf1*, *orf3* and *orf4* of the *pig* cluster, and it is therefore concluded that these genes are essential for pigment biosynthesis in this species. However, the main drawback of transposon mutagenesis is that it causes gross disruption in the mutated region and the possibility of polar effects on downstream genes is very likely, particularly if they are translationally coupled

with upstream genes. As sequence data are now available, a more directed mutagenesis approach should be taken in the future to create in-frame insertional mutants, for example. Ideally, every ORF should be mutated. Future work should also involve further attempts to characterise orange-pigmented and hyperpigmented mutants which were isolated during this study. It is likely that a regulator-encoding gene(s) has been inactivated in hyperpigmented strains. It is predicted that orange mutants are defective in O-methyltransferase.

Also identified during this study were genes which are essential for prodigiosin biosynthesis but which do not map to the *pig* cluster: a locus with homology to the *ompB* locus of *E. coli* was identified, and also a homologue of hscA which encodes a cold-shock chaperonin. It is possible that OmpB proteins are indirectly involved in pigment biosynthesis, since osmolarity regulation and other OmpR/EnvZ regulated processes are linked with cell survival in stationary phase, and secondary metabolites are also produced almost exclusively in late stages of growth. It is possible that a HscA homologue in *S. marcescens* is required for correct folding of certain Pig proteins, as it has been suggested to have a non-cold shock role in other systems. Only three such proteins have so far been identified in other bacteria, so this is a very interesting finding. Future work in further investigation of these loci should include initially, further sequencing in order to confirm these findings, and to allow accurate comparisons of gene organisation and amino acid sequences with homologues. After this, it would be necessary to express the putative proteins and also to create gene fusion strains.

Construction of gene fusion strains in *orf1* and *orf4* of the *pig* cluster allowed some preliminary work to be done to assay gene activity. It was found that *orf1* has a greater level of expression that *orf4*, and the LacZ expression profiles obtained from these experiments under conditions suitable for prodigiosin biosynthesis showed that expression patterns also differ for *orf1* and *orf4*. This suggests that they are differentially regulated. Although both appear to be active from the onset of growth, an increase in gene activity was seen in mid-log phase

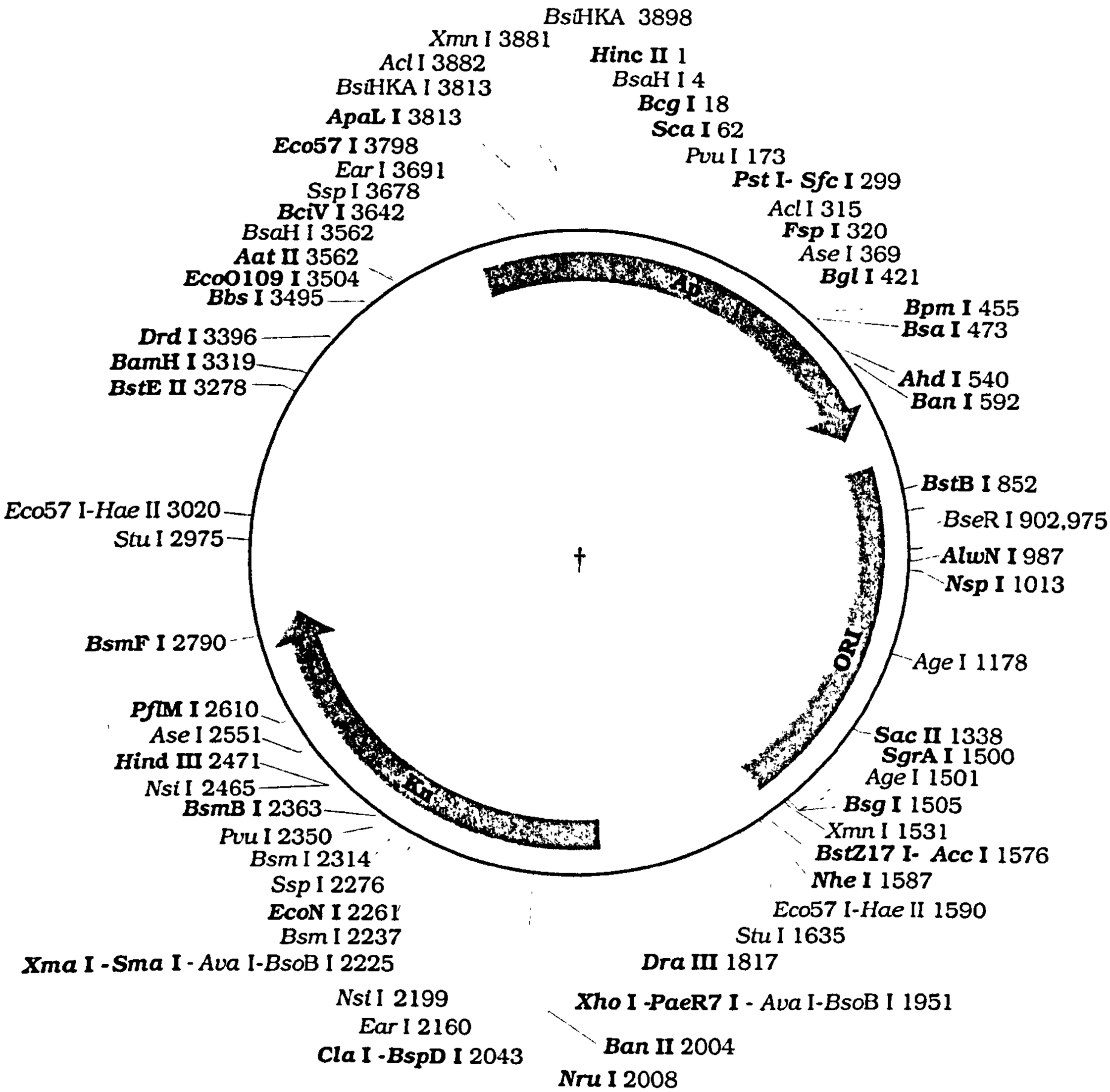
which reached a maximum in late-log phase. Crude plate assays showed that gene activity in these fusion strains is not abolished at a prodigiosin biosynthesis-restrictive temperature (37°C), so it is concluded that the encoded proteins are not temperature sensitive. It has previously been suggested that it is the enzyme that is required for the terminal condensation step of prodigiosin biosynthesis which is sensitive to increased temperature. The condensing-enzyme encoding gene is likely to be located elsewhere in the *pig* cluster. Construction of the double mutant Δ*rap*Δ*pig* strain NTSA1 suggests that activity of *orf1* locus in the *pig* cluster is Rap-dependent. Although *rap* mutation needs to be physically confirmed in this strain, this is a very interesting result. Gel retardation work is currently underway in this laboratory to assess if Rap is a DNA-binding protein, and one of the areas under analysis is the intergenic region preceding *orf1* in the *pig* cluster. Future work using gene fusion strains would of course, also involve building on the work started in this study by altering the many environmental conditions discussed previously which impact on prodigiosin biosynthesis, to assess the effects such changes have on *pig* gene activity. This may include changing pH, osmolarity, adding anions and cations to growth media and also investigating possible glucose-mediated catabolite repression. In the long-term, such studies may involve use of chemostat cultures to stringently regulate growth conditions.

Longer term future studies may involve using subcellular fractionation techniques to investigate cellular location of pig-encoded proteins. Any of the prodigiosin biosynthetic enzymes that are predicted to be membrane associated could be tested in a more directed way, now that sequence data are available, by construction of *blaM* or *phoA* fusion strains and then undertaking topology mapping. This information would be important in designing a future strategy for overexpression and purification of individual Pig proteins. A long-term aim here would be to overexpress the genes encoding putative pigment biosynthetic enzymes for *in vitro* studies on the chemistry of various biotransformations. This would require analytical and synthetic chemistry on intermediates, and if pursued,

such investigation should eventually lead to elucidation of the complete biochemistry of the pathway of prodigiosin biosynthesis in this species.

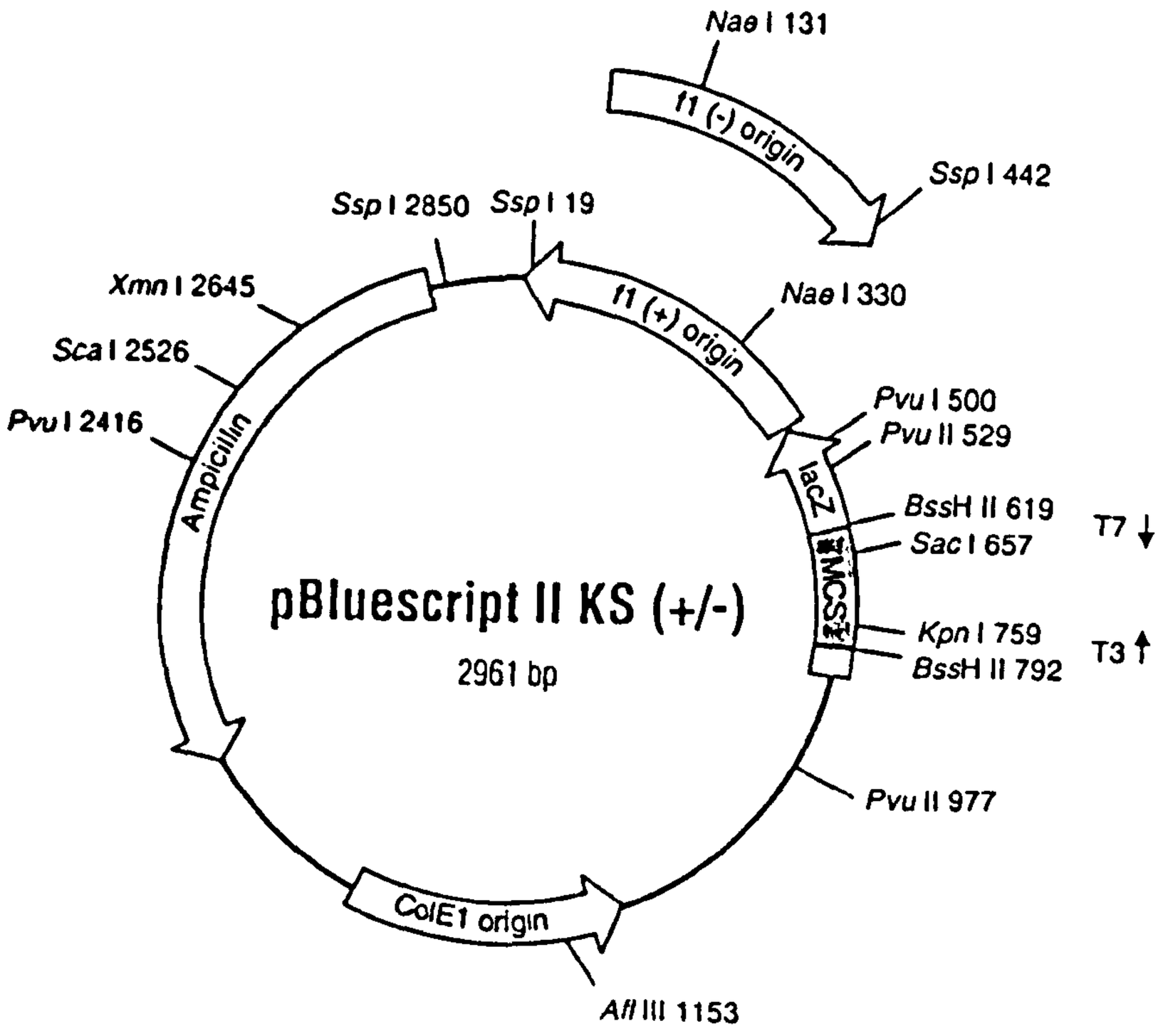
APPENDIX I

PLASMIDS, COSMIDS AND BACTERIOPHAGE VECTORS



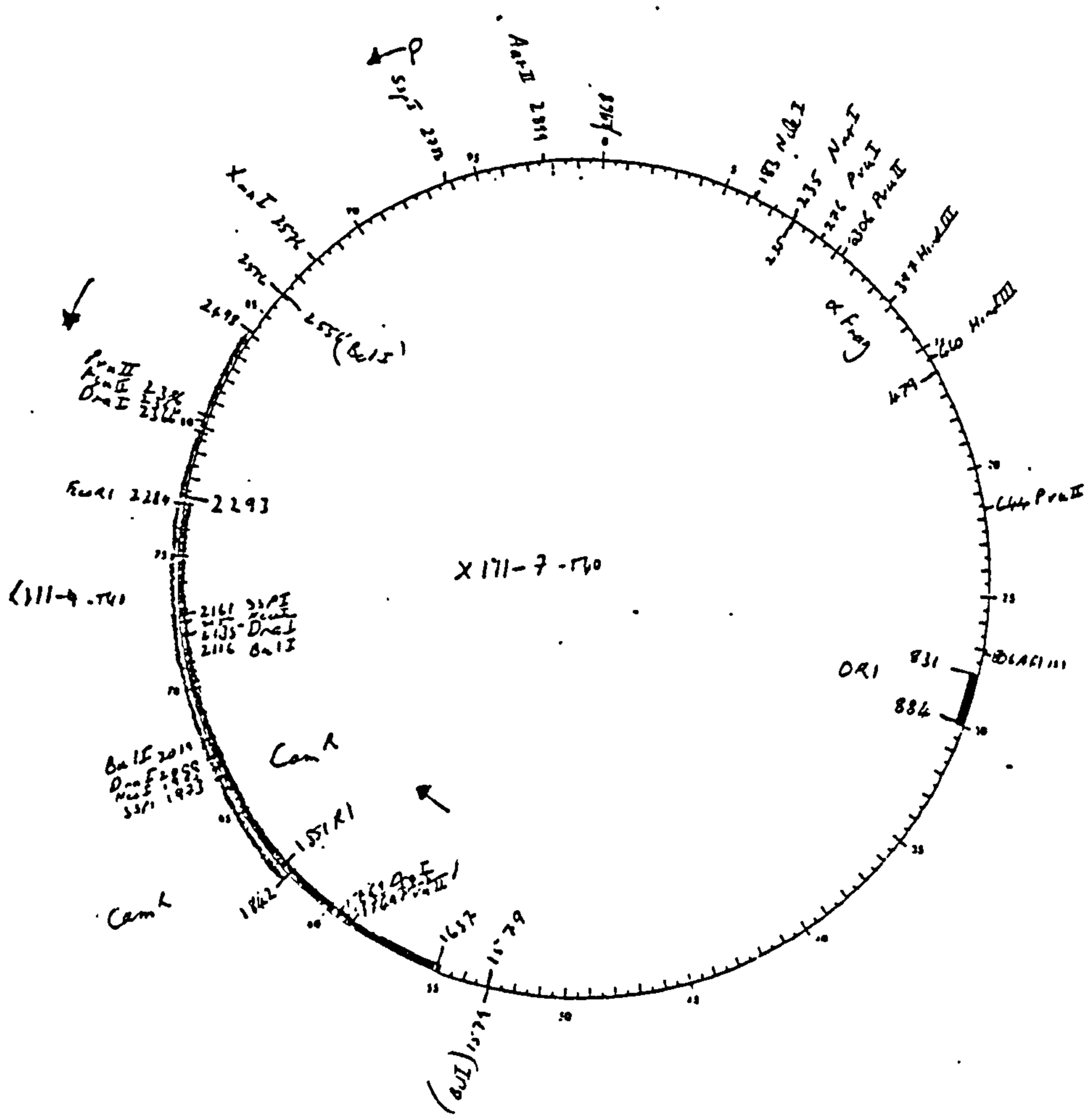
PLASMID pACYC177

pACYC177 is a low copy (copy number 10-15) cloning vector that is 3940 bp in length. It carries Kn^R and Ap^R markers (Chang and Cohen, 1978).



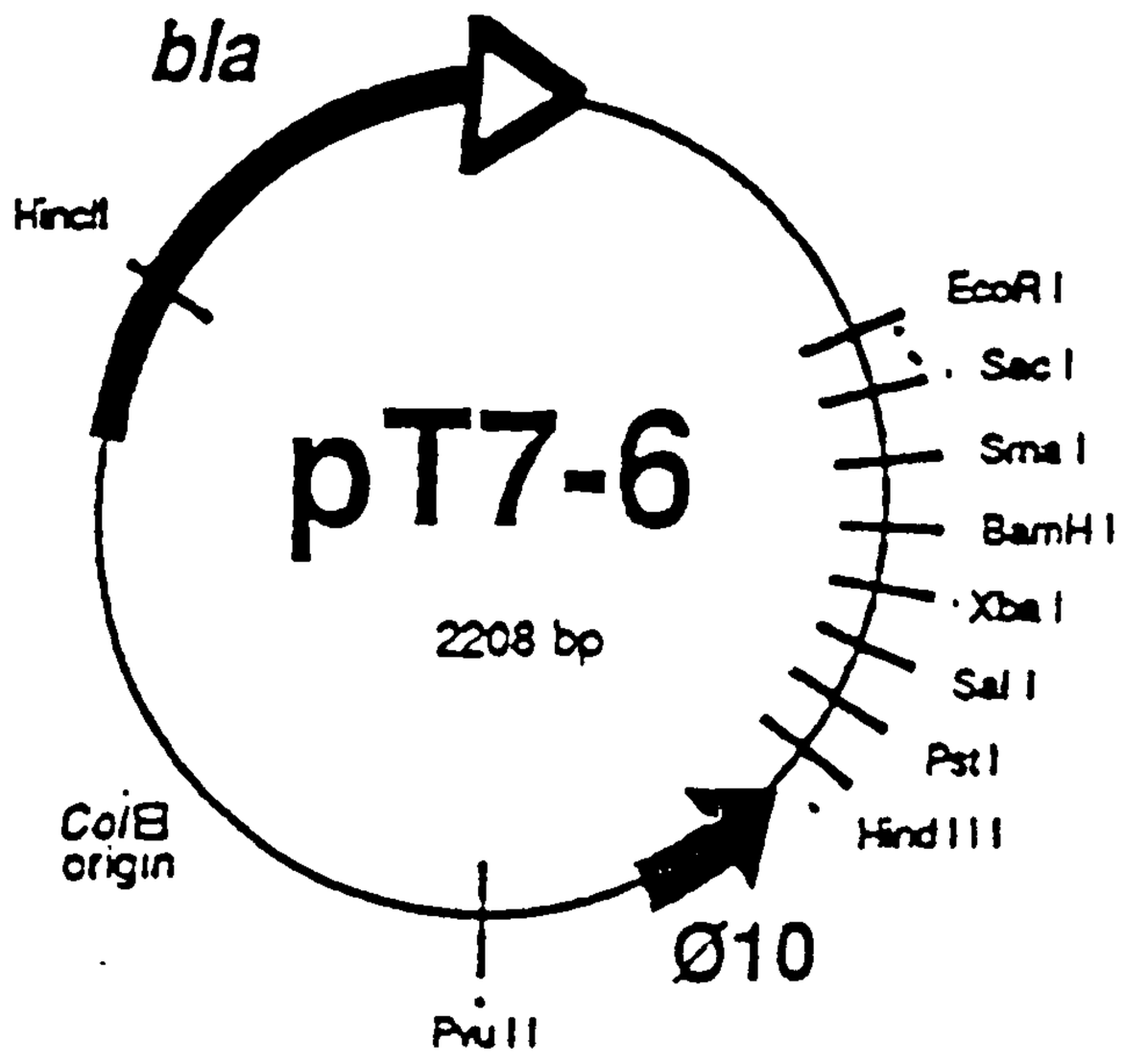
PLASMID pBLUESCRIPT II KS

pBluescript is a 2961 bp phagemid derived from pUC19. The KS designation indicates that the polylinker is orientated such *lacZ* transcription proceeds from *Kpn*I to *Sac*I. (Stratagene Ltd.)

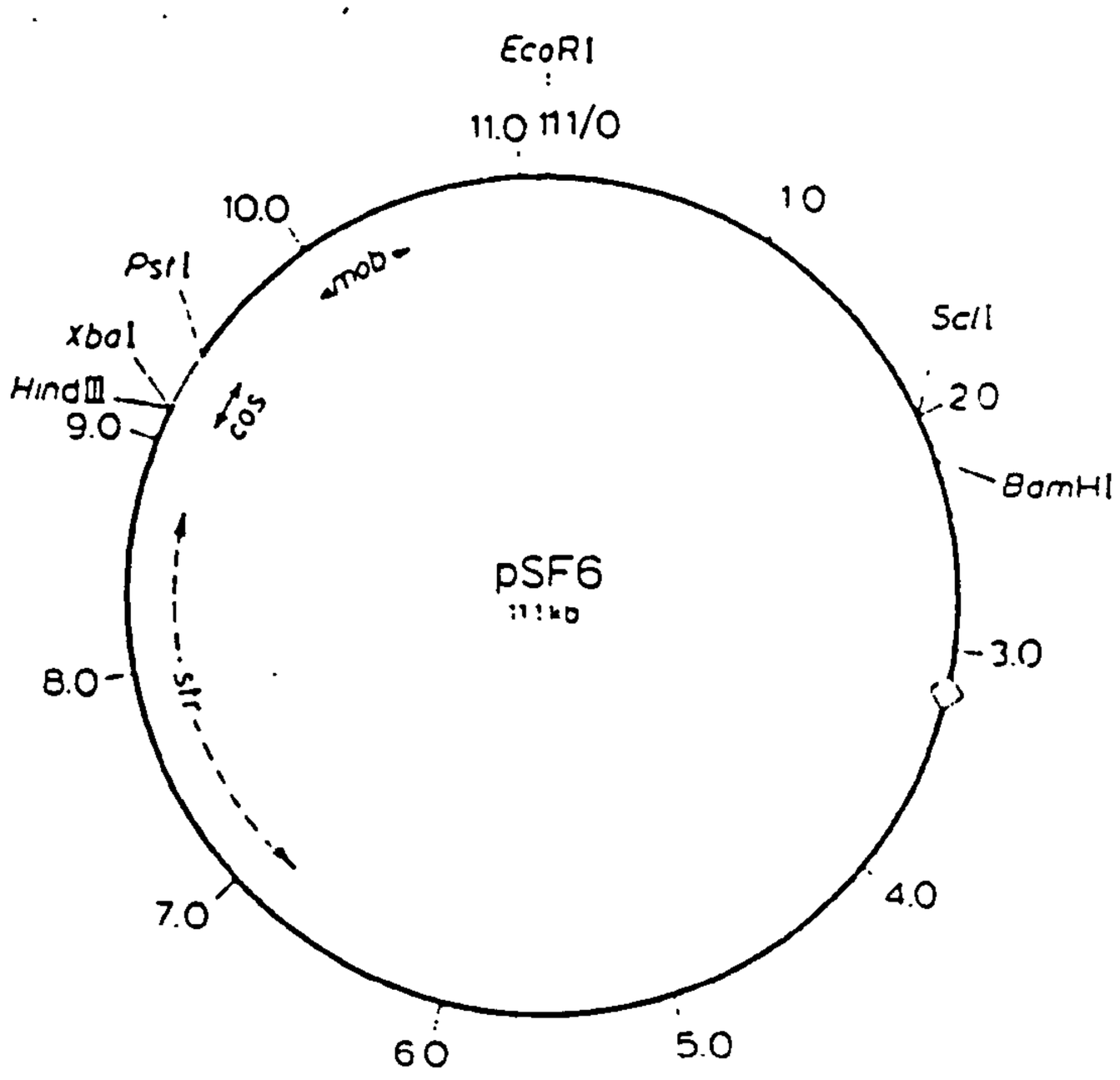


PLASMID pDAH330

pDAH330 is a very high copy number (copy number ~200) plasmid carrying a Cm^R selective marker (D. Hodgson, unpublished).

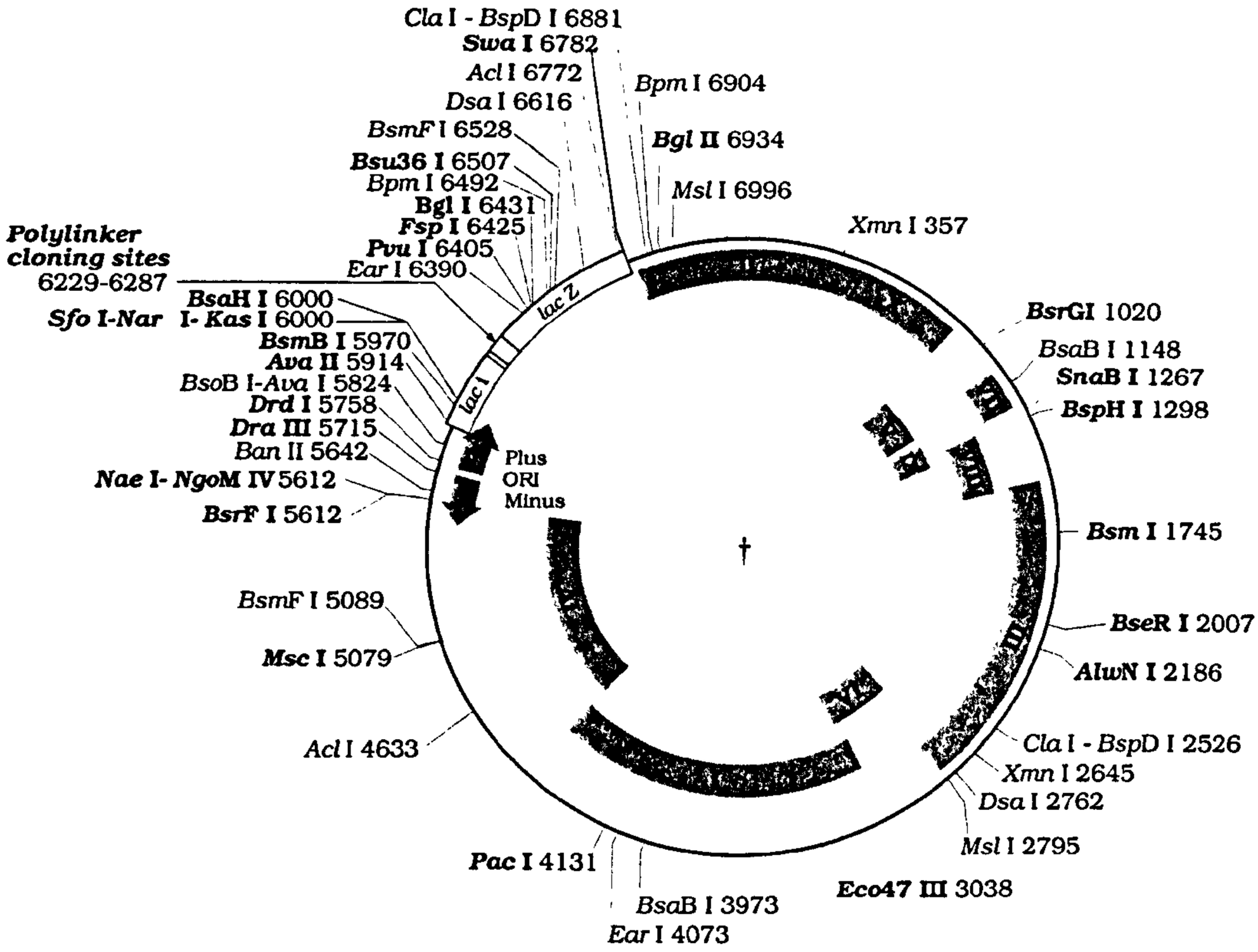


pT7-6 RNA POLYMERASE/PROMOTER VECTOR
Orientation of β-lacatamase is such that expression by T7 RNA polymerase is limited exclusively to cloned genes. (Tabor and Richardson, 1985).



COSMID pSF6

pSF6 is a low-copy cloning vector carrying the cohesive (*cos*) site of bacteriophage λ and the mobilisation (*mob*) site of the broad host-range plasmid RK2 (Selveraij *et al.*, 1984).



1

(Met) Thr Met Ile Thr Asn Ser Ser Ser Val Pro Gly Asp Pro Leu Glu Ser Thr Cys Arg His Ala Ser Leu Ala →LacZ'

6230 6250 6270 6290

Ecl136 II Sac I — Xma I Sma I — Xba I — Pst I — Hind III

atgaccatgattacgaattCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGcact

EcoR I — Kpn I Acc65 I — BamH I — Sal I Hinc II Acc I — Sph I

Ava I/BsoB I

BACTERIOPHAGE VECTOR M13 mp18

M13 is a filamentous male-specific *E. coli* bacteriophage. M13 mp18 and mp19 are both 7249 bp in length and differ only in the orientation of the 54 bp polylinker they carry (Messing and Vieira, 1982)

APPENDIX II

OLIGONUCLEOTIDE PRIMERS USED IN THIS STUDY

Appendix II. Oligonucleotide primers designed for sequencing the *pig* cluster by "primer walking"

Oligo-nucleotide reference number (ID)	Location (Fragment of *pig* cluster)[1]	Sequence 5' → 3'	Mol. weight	T_M (°C)*	Stock Conc. (μM)
3196	2.7F	CAT TGT GAA TAT TGA CC	5184.5	46	12
3543	2.7F	CTC TTG TGT CAA TAT TG	5166.4	46	26
3657	2.7F	GCT ATT CAC ATT ACG AC	5129.4	48	17
5113	2.7F	CGC TTG TGT TAT TCC TC	5118.4	50	23
3782	2.7F	GCA GCG AAT GCG GCC AC	5205.5	58	23
3195	2.6F	CCT CTT CTG AAT CAT CG	5096.4	50	22
3541	2.6F	GAT AAG TTC TCT CCT CG	5136.4	50	24
3656	2.6F	CTA CTA TCT GGC GAA TG	5185.5	50	24
3781	2.6F	GCG CAA CTC ATG GAT AC	5179.5	52	22
3193	2.0F	GGC TGA TGA CGC CAC TG	5211.5	56	23
3539	2.0F	CCT GCG CAC GAT CGA AG	5180.4	56	21
3654	2.0F	CGG CTC GAT CAG AAT CG	5195.5	54	23
3847	0.8F	GTG CGA GTG GTC ATA GC	5266.5	54	20
3194	3.2F	GCA CGC TGA TCA ATC TG	5170.4	52	23
3480	3.2F	GCT GGT GGA TTC GTT AG	5272.5	52	27
3600	3.2F	GCC TTG ATG GGC GAT GC	5242.5	56	18
3742	3.2F	GAA TCG ACA AGC ACT TC	5163.5	50	19
3814	3.2F	GTC CTG GCT GGA TAT TC	5192.4	52	21
3481	3.2R	GTT CAG TGA CTC ACC AC	5130.4	52	20
3198	3.2R	GAT ATC TCT GGG CAG TG	5241.5	52	27
3601	3.2R	GTT CAG CGC CAT CCT CG	5122.4	56	22
3743	3.2R	GCA TGA CGA ATG GTG TC	5250.5	52	21
4217	3.2R	GTA GTT CGC CGA CAT CG	5186.4	54	23
3815	3.2R	CTG GCT AAC AAC ATC TG	5154.4	50	23
4516	3.2R	GAG AGA GTC GAT TCC AG	5259.5	52	26
8065	3.2R	CGA CGA ACG GAC ATC AC	5173.5	54	23
3197	2.0R	CAG CAT GAT GCA ACG GT	5219.5	52	24
3540	2.0R	CTT GAT ATG GCG TTC AC	5176.4	50	20
3483	2.0R	GCT TGC GCC AGC GCA TC	5147.4	58	18
3655	2.0R	CCG TCA CAA TCT CTC CG	5066.4	54	26
4515	2.0R	GTT AGC CGT TTT GGC CG	5290.5	54	22
4214	2.6R	GTG CGG TAC GGT CTT GC	5233.5	56	19
3542	2.6R	CGT CAG CAT GAT CGT CC	5146.4	54	17
3813	2.6R	GTG GAT ACT GCA TTC TC	5176.4	50	17
4517	2.6R	GTT TAC CCA GTC TGG CG	5177.4	54	20
5064	2.6R	CAC AAT TTC ACC TAA GC	5098.4	48	24
3200	2.7R	CCG ATA CGC TGT GAT TG	5201.6	52	23

(Table continued...)

Oligo-nucleotide reference number (ID)	Location (Fragment of *pig* cluster)[1]	Sequence 5' → 3'	MW (Da)	T_M (°C)[2]	Stock Conc. (μM)
3544	2.7R	CTT CTG TGA TAT TCG CG	5167.4	50	24
3812	2.7R	CAG ATC TTG CAT AAG CG	5194.5	50	20
4215	2.7R	GTT GTT CTG ACT TAC GG	5207.5	50	25
4518	2.7R	GCG TGA GTA ACC GAT AG	5259.5	52	23

[1]Fragment of prodigiosin biosynthetic cluster: numbers represent the approximate length of the fragment in Kb pairs. 2.7 denotes the length of biosynthetic cluster DNA from *Pst*I / *Sma*I to *Kpn*I. Similarly, 2.6 denotes the fragment of DNA from *Kpn*I to *Bam*HI. All other fragments are *Bam*HI fragments in the 5' end of the *pig* cluster. An "F" suffix indicates a "forward" primer designed from the non-coding strand; "R" suffix indicates a "reverse" primer designed from the coding strand.
[2]Annealing temperatures for primers were calculated according to the Wallace rule (Ikatura *et. al.*, 1984)

PRIMERS USED FOR PCR AMPLIFICATION OF *rap* (Thomson, 1996)

Cyto1: 5′ GCT AAT TAT AAG GAG TA 3′ optimum T_M=39°C
Cyto2: 5′ TTT CAA GTG CCA TA CAA 3′ optimum T_M=41°C

PRIMER USED FOR SEQUENCING FROM THE ENDS OF Tn*phoA*'-2

Tn5 Pr1: 5′ CAG GAC GCT ACT TGT G 3′ optimum T_M=48°C

APPENDIX III

COMPUTER PROGRAMS AND WORLD WIDE WEB SITES

I. UNIVERSITY OF WISCONSIN GENETICS COMPUTER GROUP (UWGCG) COMPUTER PROGRAMS USED IN THIS STUDY

GAP
This program is used to align a pair of DNA or amino acid sequences in progressive, pair-wise fashion. It has adjustable stringency and calculates percentage values for identity and similarity for a given pair of sequences.

MAP
The program can be used to find restriction endonuclease sites in a DNA sequence and translate the sequence in three forward and/or three reverse reading frames. This was a useful program in this study because it allows open reading frames and the corresponding DNA sequence to be displayed together. It was also used to find restriction sites in the sequence in relation to location of ORFs.

MOTIFS
The program can be used to search for secondary structures in an amino acid sequence (e.g., signal peptidase I cleavage recognition site) and also characteristic sequence motifs (e.g. *rho*-dependent terminators) in nucleotide sequences. MOTIFS compares the query sequence to a database of specific consensus sequences for protein secondary structures and DNA motifs.

PEPDATA
The program translates DNA sequence in a continuous fashion, into potential coding regions bounded by start and stop codons. It was used in preliminary identification of areas in amino acid sequences (translated from pig cluster DNA) in six frames which might represent ORFs.

PEPTIDESTRUCTURE
The program predicts secondary structure from primary structure, i.e., amino acids sequence. The algorithms can be chosen. In this study, the prediction criteria for hydropathy in secondary structure were those of Kyte and Doolittle (1984) and for peptide folding, those of Chou and Fassman (1978) were used.

PLOTSTRUCTURE
The program displays the secondary structure and hydropathy predictions generated by PLOTSTRUCTURE graphically as a two dimensional plot.

PILEUP
This program creates a multiple sequence alignment from a group of related (homologous) DNA or amino acid sequences. Multiple alignment is created by progressive pair-wise alignment of input sequences. The stringency threshold for alignments can be adjusted and the output is saved as a multiple sequence file (~.msf), which can be used to create BOXSHADE displays with an appropriate program. It also creates the input file for PRETTY.

PRETTY
This program displays the msf calculated by PILEUP, and optionally, it can also create a consensus sequence.

REPEAT
The programme finds direct repeats in a nucleotide sequence. The stringency and size range of repeats can be altered, and all repeats of that size or greater are displayed as short alignments.

STEMLOOP
This program finds stems (inverted repeats) within a nucleotide sequence. Stem length, minimum and maximum loop sizes, and minimum number of bonds per stem must be specified. All loops, or the only the best loops, can be displayed or written into a file.

II WORLD-WIDE-WEB (WWW) SITES AND PROGRAMS USED IN THIS STUDY

BASIC LOCAL ALIGNMENT SEARCH TOOL (BLAST)
All DNA and translated amino acid sequences generated during the study were screened against non-redundant databases.
WWW sites:
www.blast.genome.ad.jp/SIT/BLAST.html (Server at Tokyo University, Japan)
www.ncbi.nih.gov/BLAST/ (Server at the National Center for Biotechnology Information, USA)

COMPUTE pI / MOLECULAR WEIGHT
This tool allows computation of the theoretical isoelectric point (pI) and molecular weight for a list of SWISS-PROT and/or TrEMBL entries, or for a user-entered amino acid sequence.
WWW site:
expasy.hcuge.ch/ch2d/pi_tool.html

ENTREZ BROWSER and GENBANK
The Entrez Browser is provided by the National Center for Biotechnology Information (NCBI), USA, which also builds, maintains and distributes the Genbank database. Genbank is the National Institute of Health (NIH) genetic sequence database, an annotated collection of publicly accessible DNA sequences. Genbank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA Databank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and Genbank at NCBI.
WWW site:
www.ncbi.nlm.nih.gov/Entrez/index.html.

PROTSCALE

Protscale allows the user to compute and represent the profile produced by any amino acid scale on a selected protein which is user-provided. An amino acid scale is defined by the numerical value assigned to each amino acid. The most frequently used scales are the hydrophobicity or hydrophilicity scales, and the secondary structure conformational parameter scales, but many other scales exist which are based on the chemical and physical properties of amino acids. This program provides 50 pre-defined scales from the literature. To calculate hydrophobicity in this study, the scale used was that based on the criteria of Kyte and Doolittle (1982)
WWW site:
expasy.hcuge.ch/cgi-bin/protscale.pl

SWISS-PROTEIN and TrEMBL

SWISS-PROT is a curated protein sequence database which provides a high level of annotations, e.g. the description of function of a database protein, domain structure, post-translational modification, variants, etc. SWISS-PROT has minimum redundancy and high level of integration with other databases.
TrEMBL is a supplementary database that contains all the translations of EMBL nucleotide sequence entries not yet integrated in SWISS-PROT.
WWW sites:
expasy.hcuge.ch/sprot/ (Home page)
expasy.hcuge.ch/sprot/sp-docu.html (Documentation for SWISS-PROT database entries)
expasy.hcuge.ch/ (Main page of the Swiss Institute for Bioinformatics)

APPENDIX IV

THE COMPLETE PRODIGIOSIN BIOSYNTHETIC CLUSTER DNA SEQUENCE AND PUTATIVE PROTEINS ENCODED BY THE 5' END OF THE CLUSTER

IV(i) THE DNA SEQUENCE OF THE *pig* CLUSTER

```
      CTGCAGAATGGGTGACGGGTAATGAGATGCTGGTCCCTCTGGCCGGCCCTGCTAACCGTC
    1 ---------+---------+---------+---------+---------+---------+ 60
      GACGTCTTACCCACTGCCCATTACTCTACGACCAGGGAGACCGGCCGGGACGATTGGCAG

      AAGGGCGTATCGCCGCCGAGAACATGCTGGGTGGTGAGCGTCGCTACCAGCGCAGTCAGG
   61 ---------+---------+---------+---------+---------+---------+ 120
      TTCCCGCATAGCGGCGGCTCTTGTACGACCCACCACTCGCAGCGATGGTCGCGTCAGTCC

      GTACGGCTATCTGTAAAGTGTTTGACCATGCTGTTGGCAGTGTGGGATTGAATGAAAAAG
  121 ---------+---------+---------+---------+---------+---------+ 180
      CATGCCGATAGACATTTCACAAACTGGTACGACAACCGTCACACCCTAACTTACTTTTTC

      CGCTGGGGCGTTTACAAATGGACTACCAGAAGGTGTATATCCATACACCCAATCATGCGA
  181 ---------+---------+---------+---------+---------+---------+ 240
      GCGACCCCGCAAATGTTTACCTGATGGTCTTCCACATATAGGTATGTGGGTTAGTACGCT

      GTTATTATCCTGGCAGTTCTCCGATTTCATTCAAGTTGCTGTTTAATCCCATTACTGGCG
  241 ---------+---------+---------+---------+---------+---------+ 300
      CAATAATAGGACCGTCAAGAGGCTAAAGTAAGTTCAACGACAAATTAGGGTAATGACCGC

      ATATTTTCGGGGCGCAGGCGGTGGGTAAGGCTGGCATTGATAAATGTATTGATGTGATAG
  301 ---------+---------+---------+---------+---------+---------+ 360
      TATAAAAGCCCCGCGTCCGCCACCCATTCCGACCGTAACTATTTACATAACTACACTATC

      CGGTGGCACAGCGGGCACATCTGAAAGTTCAGGATTTGGAATACCTTGAGCTTGCTTATG
  361 ---------+---------+---------+---------+---------+---------+ 420
      GCCACCGTGTCGCCCGTGTAGACTTTCAAGTCCTAAACCTTATGGAACTCGAACGAATAC

      CGCCTCCTTTTAATAGTGCGCGTGATGTGGTTAACCAGGCCGGTATGCTGGCGAGCAACG
  421 ---------+---------+---------+---------+---------+---------+ 480
      GCGGAGGAAAATTATCACGCGCACTACACCAATTGGTCCGGCCATACGACCGCTCGTTGC

      TGATTAATGGTGATACCAAGATTTGCCATGTCGTCGACATTGTGAATATTGACCATCAGG
  481 ---------+---------+---------+---------+---------+---------+ 540
      ACTAATTACCACTATGGTTCTAAACGGTACAGCAGCTGTAACACTTATAACTGGTAGTCC

      GACAGTGCCTGCTGGATATACGTACCTCGCAAGAGTTGAAAACGATTGGTACATATCCGG
  541 ---------+---------+---------+---------+---------+---------+ 600
      CTGTCACGGACGACCTATATGCATGGAGCGTTCTCAACTTTTGCTAACCATGTATAGGCC

      AAGCATTACATATCCCGGTTGATGAATTACGCGGTCGCTTAAATGAATTACCGAAAGACA
  601 ---------+---------+---------+---------+---------+---------+ 660
      TTCGTAATGTATAGGGCCAACTACTTAATGCGCCAGCGAATTTACTTAATGGCTTTCTGT

      AGGAAATTCTAATCGGTTGTCAATCGGGTTTGCGCGGTCATGTTGCCTATCGGTTACTCA
  661 ---------+---------+---------+---------+---------+---------+ 720
      TCCTTTAAGATTAGCCAACAGTTAGCCCAAACGCGCCAGTACAACGGATAGCCAATGAGT

      CGCAACATGGCTTCCGCGCCCGGAATTTGTCCGGGGGATATAAAACTTACAGCCTGACGT
  721 ---------+---------+---------+---------+---------+---------+ 780
      GCGTTGTACCGAAGGCGCGGGCCTTAAACAGGCCCCCTATATTTTGAATGTCGGACTGCA

      TGTCCGCCTGAACCATTAGCGTTCGGATTAATCATTGGCATGATGGCAAGTATAGGGGCC
  781 ---------+---------+---------+---------+---------+---------+ 840
      ACAGGCGGACTTGGTAATCGCAAGCCTAATTAGTAACCGTACTACCGTTCATATCCCCGG

      GTCATGCCAAACCAGATGATGTAATATAAAAATACGATTTCACTATGTGAGTGCAATAAC
  841 ---------+---------+---------+---------+---------+---------+ 900
      CAGTACGGTTTGGTCTACTACATTATATTTTTATGCTAAAGTGATACACTCACGTTATTG
```

```
     CGGTATCCTTTCGCTTGCCGCCATCTTGCAAATAAAGTTATCCAGGACATCTATATTCAT
 901 ---------+---------+---------+---------+---------+---------+ 960
     GCCATAGGAAAGCGAACGGCGGTAGAACGTTTATTTCAATAGGTCCTGTAGATATAAGTA

      TATTATGTAATATGCTTATTTATTCAATCTTTCGAGCATTGAGTAATTACTCTTGTGTCA
 961 ---------+---------+---------+---------+---------+---------+ 1020
     ATAATACATTATACGAATAAATAAGTTAGAAAGCTCGTAACTCATTAATGAGAACACAGT

     ATATTGTATTTTATTATTAGATGAGGTTATTTATTTTATTGTTCTAACTTCGGAATTTAT
1021 ---------+---------+---------+---------+---------+---------+ 1080
     TATAACATAAAATAATAATCTACTCCAATAAATAAAATAACAAGATTGAAGCCTTAAATA

     TGAACTATAAGTTTAATGATATTGTAATATTTAAATAATATATGGCAGTAAGGTTTTGTA
1081 ---------+---------+---------+---------+---------+---------+ 1140
     ACTTGATATTCAAATTACTATAACATTATAAATTTATTATATACCGTCATTCCAAAACAT

     TAAAACACAATATTAATTCAACTGATCAGTTAACCCTTTATTCAGGAAAAAACTCACGCA
1141 ---------+---------+---------+---------+---------+---------+ 1200
     ATTTTGTGTTATAATTAAGTTGACTAGTCAATTGGGAAATAAGTCCTTTTTTGAGTGCGT

     TAGGAACACCATAATCTATAAATTTATATCATGATATGGTCACGGTTTGCCTCGATACAC
1201 ---------+---------+---------+---------+---------+---------+ 1260
     ATCCTTGTGGTATTAGATATTTAAATATAGTACTATACCAGTGCCAAACGGAGCTATGTG

     CAATTCTTCTTCACCAAATTGTAACAATAAGCTTAATGAATCCCTAGCTGATGACCTTCG
1261 ---------+---------+---------+---------+---------+---------+ 1320
     GTTAAGAAGAAGTGGTTTAACATTGTTATTCGAATTACTTAGGGATCGACTACTGGAAGC

     GTCATCAGTGTCGTTTTAGCGTCGGTTTAACAATGGCCTCTATTTACTTGCAATAAAAAT
1321 ---------+---------+---------+---------+---------+---------+ 1380
     CAGTAGTCACAGCAAAATCGCAGCCAAATTGTTACCGGAGATAAATGAACGTTATTTTTA

     ACTTAGCTAAATCAATATTAAAAGTGCATCAATATTCTCACTGCTCCGTAAGTCAGAACA
1381 ---------+---------+---------+---------+---------+---------+ 1440
     TGAATCGATTTAGTTATAATTTTCACGTAGTTATAAGAGTGACGAGGCATTCAGTCTTGT

     ACGTCAAACGCTTATGCAAGATCTGGATTGTTTAATGTTTGTTTATTAAAAGTGTAATTT
1441 ---------+---------+---------+---------+---------+---------+ 1500
     TGCAGTTTGCGAATACGTTCTAGACCTAACAAATTACAAACAAATAATTTTCACATTAAA

     GTAAAATATTGTGTATTGTTGTTTTATTAAAAATTTATTATATATCAAAGGTTAAACCAT
1501 ---------+---------+---------+---------+---------+---------+ 1560
     CATTTTATAACACATAACAACAAAATAATTTTTAAATAATATATAGTTTCCAATTTGGTA

     GTGTTAATTGTGGGTATGTAACCATGTTACTGGTAACTGGAAAGCTATTCACATTACGAC
1561 ---------+---------+---------+---------+---------+---------+ 1620
     CACAATTAACACCCATACATTGGTACAATGACCATTGACCTTTCGATAAGTGTAATGCTG

     GTATTTTCCAATGTTGCATTTTGTGCTCATCAACATTAAAGATAATAGCGAAGTCAATTA
1621 ---------+---------+---------+---------+---------+---------+ 1680
     CATAAAAGGTTACAACGTAAAACACGAGTAGTTGTAATTTCTATTATCGCTTCAGTTAAT

     ACCGTTGTAGCAATGGAAGCAATGGAGTGTTTTATGGATTTTAACCTGTCAAATAGTCAG
1681 ---------+---------+---------+---------+---------+---------+ 1740
     TGGCAACATCGTTACCTTCGTTACCTCACAAAATACCTAAAATTGGACAGTTTATCAGTC

     TCAGATATTTATGAGTCAGCGTATAGATTTGCTTGCGATGTATTAGATCAAGATGCACAA
1741 ---------+---------+---------+---------+---------+---------+ 1800
     AGTCTATAAATACTCAGTCGCATATCTAAACGAACGCTACATAATCTAGTTCTACGTGTT

     ACGCGAATATCACAGAAGATTTTAAGTACTGAATTATGGAAAAAGGCCGCTGCTTATGGA
1801 ---------+---------+---------+---------+---------+---------+ 1860
     TGCGCTTATAGTGTCTTCTAAAATTCATGACTTAATACCTTTTTCCGGCGACGAATACCT
```

```
     TTTGCACATGGTCCGGTTTCTCACCAATTTGGCGGCTCGGAGCTGGGGGCGCTTGATACT
1861 ---------+---------+---------+---------+---------+---------+ 1920
     AAACGTGTACCAGGCCAAAGAGTGGTTAAACCGCCGAGCCTCGACCCCCGCGAACTATGA

     GCATTGATGATCGAGGCCCTGGGAAAAGGAAGCCGTGATATTGGATTATCGTTTTCATTA
1921 ---------+---------+---------+---------+---------+---------+ 1980
     CGTAACTACTAGCTCCGGGACCCTTTTCCTTCGGCACTATAACCTAATAGCAAAAGTAAT

     TGCGCCCATTTATGCGCTTGTGTTATTCCTCTCTATCGTTTTGGTTCAAGTGAATTGAAA
1981 ---------+---------+---------+---------+---------+---------+ 2040
     ACGCGGGTAAATACGCGAACACAATAAGGAGAGATAGCAAAACCAAGTTCACTTAACTTT

     GATAAATATCTTGAATCGTTAGTGACCGGGAAATTAATTGCAGCGAATGCGGCCACCGAA
2041 ---------+---------+---------+---------+---------+---------+ 2100
     CTATTTATAGAACTTAGCAATCACTGGCCCTTTAATTAACGTCGCTTACGCCGGTGGCTT

     CCTGATGCAGGTTCCGATATTTACAATATGCAGGCGACAGCCCAACCCTGCGAGGGGGGA
2101 ---------+---------+---------+---------+---------+---------+ 2160
     GGACTACGTCCAAGGCTATAAATGTTATACGTCCGCTGTCGGGTTGGGACGCTCCCCCCT

     TATATTCTTAATGGGAAAAAGATTTTTATTACCAATGCGCCGATTGCCGATGTATTTATC
2161 ---------+---------+---------+---------+---------+---------+ 2220
     ATATAAGAATTACCCTTTTTCTAAAAATAATGGTTACGCGGCTAACGGCTACATAAATAG

     ATCTATGCGAAAACAAATCCTGATCACGGGTTCCTGGGCGTGTCGGCCTTTCTTATCGAG
2221 ---------+---------+---------+---------+---------+---------+ 2280
     TAGATACGCTTTTGTTTAGGACTAGTGCCCAAGGACCCGCACAGCCGGAAAGAATAGCTC

     AAAGGCACGCCGGGCCTGAACGTAGGGGAGGTGATCCCGAAAGATTGTCTTTCTAATTGC
2281 ---------+---------+---------+---------+---------+---------+ 2340
     TTTCCGTGCGGCCCGGACTTGCATCCCCTCCACTAGGGCTTTCTAACAGAAAGATTAACG

     CCCTGGAGTGAGATCGTTTTCAACGACATTTTTATTCCTCAATCACAGCGTATCGGGATG
2341 ---------+---------+---------+---------+---------+---------+ 2400
     GGGACCTCACTCTAGCAAAAGTTGCTGTAAAAATAAGGAGTTAGTGTCGCATAGCCCTAC

     GAAGGTGCGGGCGGGGCTATTTTCCATGATTCAATGATTTGGGAAAAAGGCTGTTTGTCG
2401 ---------+---------+---------+---------+---------+---------+ 2460
     CTTCCACGCCCGCCCCGATAAAAGGTACTAAGTTACTAAACCCTTTTTCCGACAAACAGC

     GCCCTGTTTGTGGGGGGATTGGCGCGCCTTTTAGAAACGACCCTAGAGTATGCCAAAGCG
2461 ---------+---------+---------+---------+---------+---------+ 2520
     CGGGACAAACACCCCCCTAACCGCGCGGAAAATCTTTGCTGGGATCTCATACGGTTTCGC

     CGCCAGCAATTTGGTAAGGCGATCGGTCAGTTCCAATCTGTCTCTAATCGAATTATCGAT
2521 ---------+---------+---------+---------+---------+---------+ 2580
     GCGGTCGTTAAACCATTCCGCTAGCCAGTCAAGGTTAGACAGAGATTAGCTTAATAGCTA

     ATGAAACTGCGCCTGGAGCAGTGTCGGTTGATGCTCTATCGCGCCTGTTGGAAACATGAT
2581 ---------+---------+---------+---------+---------+---------+ 2640
     TACTTTGACGCGGACCTCGTCACAGCCAACTACGAGATAGCGCGGACAACCTTTGTACTA

     CAGGGTCAGGATGCCGAAGCCGATATCGCCATGAGTAAACTGCTGATTTCCGAGTACGCG
2641 ---------+---------+---------+---------+---------+---------+ 2700
     GTCCCAGTCCTACGGCTTCGGCTATAGCGGTACTCATTTGACGACTAAAGGCTCATGCGC

     GTTCAATCCGGTCTGGATGCCATTCAAACTTTTGGCGGAGCAGCGATGGATCAGGAACTC
2701 ---------+---------+---------+---------+---------+---------+ 2760
     CAAGTTAGGCCAGACCTACGGTAAGTTTGAAAACCGCCTCGTCGCTACCTAGTCCTTGAG

     GGCCTGGTTCGCCATCTTTTGAATATGATCCCGAGTCGAATTTTCTCCGGTACCAATGAT
2761 ---------+---------+---------+---------+---------+---------+ 2820
     CCGGACCAAGCGGTAGAAAACTTATACTAGGGCTCAGCTTAAAAGAGGCCATGGTTACTA
```

```
     ATCCAGAAAGAGATTATTGCCCGTAAACTCGGTTTAAGGGGAACATCATCATGATTATTC
2821 ---------+---------+---------+---------+---------+---------+ 2880
     TAGGTCTTTCTCTAATAACGGGCATTTGAGCCAAATTCCCCTTGTAGTAGTACTAATAAG

     AACGGCTCTTCGGCATCCTCTATATGCTTGCCGGTTTGGCGAAAGCTTTCCCGCAGTTTG
2881 ---------+---------+---------+---------+---------+---------+ 2940
     TTGCCGAGAAGCCGTAGGAGATATACGAACGGCCAAACCGCTTTCGAAAGGGCGTCAAAC

     AGAATGTTCCCGCTGTTCTTCGGCAGGCGGCAATTGCCAACCAGGGAACCTGGTACGCGG
2941 ---------+---------+---------+---------+---------+---------+ 3000
     TCTTACAAGGGCGACAAGAAGCCGTCCGCCGTTAACGGTTGGTCCCTTGGACCATGCGCC

     CGGCGAGTATCTGGCTGGGCGCACATGGTGATGTGATCAATATCCTGGTGGGAGTGGTGC
3001 ---------+---------+---------+---------+---------+---------+ 3060
     GCCGCTCATAGACCGACCCGCGTGTACCACTACACTAGTTATAGGACCACCCTCACCACG

     TGTTTGGGTCGGGAGTGATATTAATGCTCAACCCACTCTGGACGACGCTGGTGATCTACG
3061 ---------+---------+---------+---------+---------+---------+ 3120
     ACAAACCCAGCCCTCACTATAATTACGAGTTGGGTGAGACCTGCTGCGACCACTAGATGC

     CCCAGCTTTTGATGATGGCTGTCTTTGTGGTGATTTTGCATCAGTCCCAGCCCCAGGTCA
3121 ---------+---------+---------+---------+---------+---------+ 3180
     GGGTCGAAAACTACTACCGACAGAAACACCACTAAAACGTAGTCAGGGTCGGGGTCCAGT

     TGTTGCTGGATGGCGTATTTGCGCTGGCCGCGCTTTACATGCTGCGCGGTCAGTATCACC
3181 ---------+---------+---------+---------+---------+---------+ 3240
     ACAACGACCTACCGCATAAACGCGACCGGCGCGAAATGTACGACGCGCCAGTCATAGTGG

     GTAAGCCTAAGCCGAGAACCTTCCCGACCACGTCTTTTTCGCTGCCCACACCCTCTTCTG
3241 ---------+---------+---------+---------+---------+---------+ 3300
     CATTCGGATTCGGCTCTTGGAAGGGCTGGTGCAGAAAAAGCGACGGGTGTGGGAGAAGAC

     AATCATCGTTTTCTGCTCCCTTGGGTGATGAGTATGATGTCGTGATTATCGGCGGTGGTG
3301 ---------+---------+---------+---------+---------+---------+ 3360
     TTAGTAGCAAAAGACGAGGGAACCCACTACTCATACTACAGCACTAATAGCCGCCACCAC

     CCTCCGGACTGACGGCGGCCAGTGAATTTACTCATGAACGGGTGTTGGTGCTTGAAAAAA
3361 ---------+---------+---------+---------+---------+---------+ 3420
     GGAGGCCTGACTGCCGCCGGTCACTTAAATGAGTACTTGCCCACAACCACGAACTTTTTT

     GCTCCACTTTCGGGGGGAATGCCCGCTACCACACCTTTAACCGGTTAAAGCATCCCACCG
3421 ---------+---------+---------+---------+---------+---------+ 3480
     CGAGGTGAAAGCCCCCCTTACGGGCGATGGTGTGGAAATTGGCCAATTTCGTAGGGTGGC

     CCGGTGTTTGTTTTCAAGAACCGTTTCCCGGTTCGAATATGCTGCGCTTACTGAAAAAGA
3481 ---------+---------+---------+---------+---------+---------+ 3540
     GGCCACAAACAAAAGTTCTTGGCAAAGGGCCAAGCTTATACGACGCGAATGACTTTTTCT

     TTGGTCTGGAGGGAAAATACAAGTCCAACGAAAAGGACACGCTGGTCTTTTTTGATACCT
3541 ---------+---------+---------+---------+---------+---------+ 3600
     AACCAGACCTCCCTTTTATGTTCAGGTTGCTTTTCCTGTGCGACCAGAAAAAACTATGGA

     TTTTATTGCTCAAATGCTTAGGTGAAATTGTGGTCGGCTTTATTAAGCAACCACGTTATC
3601 ---------+---------+---------+---------+---------+---------+ 3660
     AAAATAACGAGTTTACGAATCCACTTTAACACCAGCCGAAATAATTCGTTGGTGCAATAG

     TGCTCAAACTCTCGGTCTGGGGGCTGACCAGCCAGCTTTTTCTGCATGCGATAATTGGTA
3661 ---------+---------+---------+---------+---------+---------+ 3720
     ACGAGTTTGAGAGCCAGACCCCCGACTGGTCGGTCGAAAAAGACGTACGCTATTAACCAT

     AACCCTACGTGGTAGCGGCCAAACAACTTGGTGACCCGATCTTTGCCGATCTTTATACCT
3721 ---------+---------+---------+---------+---------+---------+ 3780
     TTGGGATGCACCATCGCCGGTTTGTTGAACCACTGGGCTAGAAACGGCTAGAAATATGGA
```

```
     TTCTGGATAAGTTCTCTCCTCGTGGCGACTTTTATCCACGCCTGCCCTGGACACCGAACG
3781 ---------+---------+---------+---------+---------+---------+ 3840
     AAGACCTATTCAAGAGAGGAGCACCGCTGAAAATAGGTGCGGACGGGACCTGTGGCTTGC

     GTTCCTGGAGCAAAGCGCATATGGAGTTGCTCGATAATATTTCCCTTTACACCTATTTGT
3841 ---------+---------+---------+---------+---------+---------+ 3900
     CAAGGACCTCGTTTCGCGTATACCTCAACGAGCTATTATAAAGGGAAATGTGGATAAACA

     TTGAGCCGGATAAGCTTGGCCGGTTACCGGAACAGCTGCGGCCACCCGCCAGACTGGGTA
3901 ---------+---------+---------+---------+---------+---------+ 3960
     AACTCGGCCTATTCGAACCGGCCAATGGCCTTGTCGACGCCGGTGGGCGGTCTGACCCAT

     AACTGGTTGAGAATGCAGTATCCACCACCTTGCGGGTTGAGTGTCTGGATATTCATGATG
3961 ---------+---------+---------+---------+---------+---------+ 4020
     TTGACCAACTCTTACGTCATAGGTGGTGGAACGCCCAACTCACAGACCTATAAGTACTAC

     TCTCTGCCTATGTCGGTTTGCACTTTCTGGTGGGCTATCTGCGCGGAAACCTGGTCACCC
4021 ---------+---------+---------+---------+---------+---------+ 4080
     AGAGACGGATACAGCCAAACGTGAAAGACCACCCGATAGACGCGCCTTTGGACCAGTGGG

     TGCCTGGCGGTAATGGCAGTATCAGTGCGGGGTTGTGTAAGTATCTAAGCCATCAGCGCA
4081 ---------+---------+---------+---------+---------+---------+ 4140
     ACGGACCGCCATTACCGTCATAGTCACGCCCCAACACATTCATAGATTCGGTAGTCGCGT

     ATGTCACGTTGCAAAACCATGTCCAGTTGACAGCGGTTGAGCCGCAGCACAATGGTACAT
4141 ---------+---------+---------+---------+---------+---------+ 4200
     TACAGTGCAACGTTTTGGTACAGGTCAACTGTCGCCAACTCGGCGTCGTGTTACCATGTA

     GCATCCAGTTCACGATCAATGGTCAACCCCGCCAAGTGCAGGCTCAACAGATCATTTGGG
4201 ---------+---------+---------+---------+---------+---------+ 4260
     CGTAGGTCAAGTGCTAGTTACCAGTTGGGGCGGTTCACGTCCGAGTTGTCTAGTAAACCC

     CCGCGCCTAAAACGCAACTTGCCACATGGCTGCCGGGATTGCCGGCCAAACAGTTGGCGG
4261 ---------+---------+---------+---------+---------+---------+ 4320
     GGCGCGGATTTTGCGTTGAACGGTGTACCGACGGCCCTAACGGCCGGTTTGTCAACCGCC

     CCATCAAGAACATTCGTCATGAAGACTACTATCTGGCGAATGTATTCCTGTCAAAACCGG
4321 ---------+---------+---------+---------+---------+---------+ 4380
     GGTAGTTCTTGTAAGCAGTACTTCTGATGATAGACCGCTTACATAAGGACAGTTTTGGCC

     TGCTGGGTCATTCGTTTGGTGGCTATATGATCGAACCGGACAGCAATAAAGATCCGTTCT
4381 ---------+---------+---------+---------+---------+---------+ 4440
     ACGACCCAGTAAGCAAACCACCGATATACTAGCTTGGCCTGTCGTTATTTCTAGGCAAGA

     CTTGGTGTAAAGCGGGGACTTGCCTGGTGGCCAACTGGATGGACGATCATGCTGACGTGG
4441 ---------+---------+---------+---------+---------+---------+ 4500
     GAACCACATTTCGCCCCTGAACGGACCACCGGTTGACCTACCTGCTAGTACGACTGCACC

     ATGTGGGCGTGCTGACATTGCTTAAACCCACGACGCGCTCAGAGCGGCAAGACCGTACCG
4501 ---------+---------+---------+---------+---------+---------+ 4560
     TACACCCGCACGACTGTAACGAATTTGGGTGCTGCGCGAGTCTCGCCGTTCTGGCATGGC

     CACAGAATGCGTTTTTAGCGCTACAACAGCAAACCTATGCCGAAATTGCCAAGGTTCTGC
4561 ---------+---------+---------+---------+---------+---------+ 4620
     GTGTCTTACGCAAAAATCGCGATGTTGTCGTTTGGATACGGCTTTAACGGTTCCAAGACG

     GCAACATAGGGATTGGTGCCGAGGTTATTGAGGATATCCAGATCTGGTATTGGCCCGCAG
4621 ---------+---------+---------+---------+---------+---------+ 4680
     CGTTGTATCCCTAACCACGGCTCCAATAACTCCTATAGGTCTAGACCATAACCGGGCGTC

     GGCTGGTGACGTCGGTAGTGGGGCAACAGGCTGAAGGTGTATTCGAAACCGCCCGTCAGT
4681 ---------+---------+---------+---------+---------+---------+ 4740
     CCGACCACTGCAGCCATCACCCCGTTGTCCGACTTCCACATAAGCTTTGGCGGGCAGTCA
```

```
     CATTTGAAAATATTCATTTCGCTAATCAGGACAGTGTCGGTGTCGGCAATATCGAGAGTG
4741 --------+---------+---------+---------+---------+---------+ 4800
     GTAAACTTTTATAAGTAAAGCGATTAGTCCTGTCACAGCCACAGCCGTTATAGCTCTCAC

     CCATCCTGTCGGGAATTGATGCAGCCAACGCGGTTAAAGCGCAACTCATGGATACGGAAA
4801 --------+---------+---------+---------+---------+---------+ 4860
     GGTAGGACAGCCCTTAACTACGTCGGTTGCGCCAATTTCGCGTTGAGTACCTATGCCTTT

     ATGTCGTGGAGGTGGCGGGATGAATCAACCACTGGTTGTTGAAATATCCGGTGATAAGGC
4861 --------+---------+---------+---------+---------+---------+ 4920
     TACAGCACCTCCACCGCCCTACTTAGTTGGTGACCAACAACTTTATAGGCCACTATTCCG

     GCTTGAACATCACCACTTGGGCGGTAAGGGTTATTCACTCAACAACCTGATTCATGCGGG
4921 --------+---------+---------+---------+---------+---------+ 4980
     CGAACTTGTAGTGGTGAACCCGCCATTCCCAATAAGTGAGTTGTTGGACTAAGTACGCCC

     CCTGCCAGTACCTTCGGCATTTTGTGTGACGGCACAAGCCTACCAACAGTTTATTGAAGA
4981 --------+---------+---------+---------+---------+---------+ 5040
     GGACGGTCATGGAAGCCGTAAAACACACTGCCGTGTTCGGATGGTTGTCAAATAACTTCT

     GGTGGTGCCGGGGGCGGAATTAACTGACGGCGACCTGATTGCGGTGCGTGACGCCATCCT
5041 --------+---------+---------+---------+---------+---------+ 5100
     CCACCACGGCCCCCGCCTTAATTGACTGCCGCTGGACTAACGCCACGCACTGCGGTAGGA

     GCATGCCGATATTCCCGACTCGCTAAAGCAGGCTATCGGTGATGCTTATCAACACCTGGG
5101 --------+---------+---------+---------+---------+---------+ 5160
     CGTACGGCTATAAGGGCTGAGCGATTTCGTCCGATAGCCACTACGAATAGTTGTGGACCC

     CCATGACACGACCATTGCGGTTCGTTCTTCGGCATTGGATGAAGATGGTCAACGTCAGTC
5161 --------+---------+---------+---------+---------+---------+ 5220
     GGTACTGTGCTGGTAACGCCAAGCAAGAAGCCGTAACCTACTTCTACCAGTTGCAGTCAG

     ATTTGCCGGGCAGTACGAGACTTACCTGCATGTGAAGGGTTCAGAGGCAGTGTTGCACAA
5221 --------+---------+---------+---------+---------+---------+ 5280
     TAAACGGCCCGTCATGCTCTGAATGGACGTACACTTCCCAAGTCTCCGTCACAACGTGTT

     GGTTCAGGCTTGTTGGGCGTCGCTTTGGGCTGAGCGAGCGGCTCAATATCGTCATGAATC
5281 --------+---------+---------+---------+---------+---------+ 5340
     CCAAGTCCGAACAACCCGCAGCGAAACCCGACTCGCTCGCCGAGTTATAGCAGTACTTAG

     TGCGTCGCACAGTGCCATTGCCGTGATTCTACAAGTGATGGTTGATGCCGATGCAGCGGG
5341 --------+---------+---------+---------+---------+---------+ 5400
     ACGCAGCGTGTCACGGTAACGGCACTAAGATGTTCACTACCAACTACGGCTACGTCGCCC

     AGTGATGTTTACTCAGGATCCTCTGTCAGGGAGCACCGATAAGGTGGTGATTGACAGTTG
5401 --------+---------+---------+---------+---------+---------+ 5460
     TCACTACAAATGAGTCCTAGGAGACAGTCCCTCGTGGCTATTCCACCACTAACTGTCAAC

     TTGGGGGCTGGGGGAAGGCGTGGTTTCCGGGCAAGTCACCACAGATAGTTTCACGCTGGA
5461 --------+---------+---------+---------+---------+---------+ 5520
     AACCCCCGACCCCCTTCCGCACCAAAGGCCCGTTCAGTGGTGTCTATCAAAGTGCGACCT

     TAAAGCCACCGGTGAGCTATGTGATCAGCAGATTCGCCACAAACCGAATTACTGCCAGCG
5521 --------+---------+---------+---------+---------+---------+ 5580
     ATTTCGGTGGCCACTCGATACACTAGTCGTCTAAGCGGTGTTTGGCTTAATGACGGTCGC

     GGACGAACATGGTCTGGTGACACTGTTGCAAACCCCGGAGGCCAAACGGGATCTCCCCAG
5581 --------+---------+---------+---------+---------+---------+ 5640
     CCTGCTTGTACCAGACCACTGTGACAACGTTTGGGGCCTCCGGTTTGCCCTAGAGGGGTC

     TCTGACGCCGGCTCAGTTGCAACAGTTGGTCACGCTGGCCAGACAGGCCCAGCTTATCTA
5641 --------+---------+---------+---------+---------+---------+ 5700
     AGACTGCGGCCGAGTCAACGTTGTCAACCAGTGCGACCGGTCTGTCCGGGTCGAATAGAT
```

```
     CAGCACCGAACTGGATATTGAATGGGCGGTGAAAGATGACAAAGTCTGGTTATTACAGGC
5701 --------+---------+---------+---------+---------+---------+ 5760
     GTCGTGGCTTGACCTATAACTTACCCGCCACTTTCTACTGTTTCAGACCAATAATGTCCG

     GCGTCCGGTAACCACATCGGCCAAAACGGCTAACGTTATCTACGCCAATCCGTGGGAGAG
5761 --------+---------+---------+---------+---------+---------+ 5820
     CGCAGGCCATTGGTGTAGCCGGTTTTGCCGATTGCAATAGATGCGGTTAGGCACCCTCTC

     CGACCCGGCTGCGAAGGAGGGCGCTTTTTTCTCGCGGATGGATACCGGAGAGATTGTGAC
5821 --------+---------+---------+---------+---------+---------+ 5880
     GCTGGGCCGACGCTTCCTCCCGCGAAAAAAGAGCGCCTACCTATGGCCTCTCTAACACTG

     GGGGCTGATGACGCCACTGGGGTTGTCATTTTGTCAGTTCTATCAAAAGCACATTCATGG
5881 --------+---------+---------+---------+---------+---------+ 5940
     CCCCGACTACTGCGGTGACCCCAACAGTAAAACAGTCAAGATAGTTTTCGTGTAAGTACC

     CCCGGCCATCAAGACCATGGGGCTTGCCGATATTAGTCACTGGCAGATCTATATGGGCTA
5941 --------+---------+---------+---------+---------+---------+ 6000
     GGGCCGGTAGTTCTGGTACCCCGAACGGCTATAATCAGTGACCGTCTAGATATACCCGAT

     TATCCAGGGCTATGTTTATCTGAATATTTCCGGTTCGGCCTATATGCTCAGGCAGTGTCC
6001 --------+---------+---------+---------+---------+---------+ 6060
     ATAGGTCCCGATACAAATAGACTTATAAAGGCCAAGCCGGATATACGAGTCCGTCACAGG

     ACCGACCCGTAATGAAATGAAGTTCACCACCCGCTATGCCACGGACGAGATCGATTTTAA
6061 --------+---------+---------+---------+---------+---------+ 6120
     TGGCTGGGCATTACTTTACTTCAAGTGGTGGGCGATACGGTGCCTGCTCTAGCTAAAATT

     GGACTATAAAAACCCCTATGGTGCAGGCGTGCAGGGCTGGGATTATGCCAAGAGTTGTTG
6121 --------+---------+---------+---------+---------+---------+ 6180
     CCTGATATTTTTGGGGATACCACGTCCGCACGTCCCGACCCTAATACGGTTCTCAACAAC

     GTACTGGCTAAAACAGCAAGTCCGTAATATGCGCAGCGCCGCCAGGACCGTCGAGCAGAT
6181 --------+---------+---------+---------+---------+---------+ 6240
     CATGACCGATTTTGTCGTTCAGGCATTATACGCGTCGCGGCGGTCCTGGCAGCTCGTCTA

     GATTGCCCTGCGTCAGGACGAAACCACACGGTTTCTGGGGCTTGATTTGACCGCCATGAC
6241 --------+---------+---------+---------+---------+---------+ 6300
     CTAACGGGACGCAGTCCTGCTTTGGTGTGCCAAAGACCCCGAACTAAACTGGCGGTACTG

     GCTTCAGCAGTTGGACCAAGAGCTGCAACGGATTGACCGCTTTTTCCTCGATAGCTGTGC
6301 --------+---------+---------+---------+---------+---------+ 6360
     CGAAGTCGTCAACCTGGTTCTCGACGTTGCCTAACTGGCGAAAAAGGAGCTATCGACACG

     GGCGTATATGCCTTTTTTCCTGCAATCATTTGCCCTTTACGATGCGCTGGCGCAAGCCTG
6361 --------+---------+---------+---------+---------+---------+ 6420
     CCGCATATACGGAAAAAAGGACGTTAGTAAACGGGAAATGCTACGCGACCGCGTTCGGAC

     TGAACGCCATATCAAGGATGGTAAAGGTCTGCAAAATCGCATCAAGGCGTCAATGAATAA
6421 --------+---------+---------+---------+---------+---------+ 6480
     ACTTGCGGTATAGTTCCTACCATTTCCAGACGTTTTAGCGTAGTTCCGCAGTTACTTATT

     CCTGCGCACGATCGAAGTGACGCTCGGTATCATCAAATTGGTTGCGACAGTTAACCAACA
6481 --------+---------+---------+---------+---------+---------+ 6540
     GGACGCGTGCTAGCTTCACTGCGAGCCATAGTAGTTTAACCAACGCTGTCAATTGGTTGT

     GACTGAATTAAAAGCGTTATTTGAACAGCATCGTGCCGATGAGTTAGTCACCTTGCTACC
6541 --------+---------+---------+---------+---------+---------+ 6600
     CTGACTTAATTTTCGCAATAAACTTGTCGTAGCACGGCTACTCAATCAGTGGAACGATGG

     TGTCCATGACATATCCCGTGCTTTCTGGCAGGGTGATTTTGAAGATTTTCTGGTTGAATT
6601 --------+---------+---------+---------+---------+---------+ 6660
     ACAGGTACTGTATAGGGCACGAAAGACCGTCCCACTAAAACTTCTAAAAGACCAACTTAA
```

```
     TGGTTCTCGCGGTCGCCAGGAGTTTGATCTGAGTATCCCGCGTTGGCGTGATGACCCGAG
6661 ---------+---------+---------+---------+---------+---------+ 6720
     ACCAAGAGCGCCAGCGGTCCTCAAACTAGACTCATAGGGCGCAACCGCACTACTGGGCTC

     TTATTTATTGCAAGTAATGAAGATGTATTTGCAACACCCGGTAGATTTGCACAAGAAGCT
6721 ---------+---------+---------+---------+---------+---------+ 6780
     AATAAATAACGTTCATTACTTCTACATAAACGTTGTGGGCCATCTAAACGTGTTCTTCGA

     CAGGGAAACGGAACTGCTGCGTCAGCAGGATAGTGAAGCACTGTTCAGTGCGATGTCCTG
6781 ---------+---------+---------+---------+---------+---------+ 6840
     GTCCCTTTGCCTTGACGACGCAGTCGTCCTATCACTTCGTGACAAGTCACGCTACAGGAC

     GTCGGGCCGTTTCAAGCTCAAAACCCTCATCAAGCTGTATGGCATGATGGCCGAACGTCG
6841 ---------+---------+---------+---------+---------+---------+ 6900
     CAGCCCGGCAAAGTTCGAGTTTTGGGAGTAGTTCGACATACCGTACTACCGGCTTGCAGC

     TGAAGCGACGCGGCCAACGTTTATCACCGAAACCTGGTTCTACCGTTGCATCATGCTGGA
6901 ---------+---------+---------+---------+---------+---------+ 6960
     ACTTCGCTGCGCCGGTTGCAAATAGTGGCTTTGGACCAAGATGGCAACGTAGTACGACCT

     AGTTCTACGGCGTCTGGACGCGCAGGGGATTGCCAGTAGTGCCGATCTGCCCTATGTCGA
6961 ---------+---------+---------+---------+---------+---------+ 7020
     TCAAGATGCCGCAGACCTGCGCGTCCCCTAACGGTCATCACGGCTAGACGGGATACAGCT

     TTTTGAACAGTTCCGTGCGTATGTGGCGGGGACCATCCCAGCTGAACAGGCGTTTTCCAA
7021 ---------+---------+---------+---------+---------+---------+ 7080
     AAAACTTGTCAAGGCACGCATACACCGCCCCTGGTAGGGTCGACTTGTCCGCAAAAGGTT

     AGCGCGGCTCGATCAGAATCGCCATCAACATTTATTCAATTTACATGCCGAAGAACCGCC
7081 ---------+---------+---------+---------+---------+---------+ 7140
     TCGCGCCGAGCTAGTCTTAGCGGTAGTTGTAAATAAGTTAAATGTACGGCTTCTTGGCGG

     GATGGCGATTGTCGGTCCCTATACCCCCAAAGTGAAAGCACCCACGCAGGATGATAAAAC
7141 ---------+---------+---------+---------+---------+---------+ 7200
     CTACCGCTAACAGCCAGGGATATGGGGGTTTCACTTTCGTGGGTGCGTCCTACTATTTTG

     CATCCGGTCACTCACTGGCCTGGCGGCCAGTCCTGGCAACGTAGTGGCCAAAGCGCGGGT
7201 ---------+---------+---------+---------+---------+---------+ 7260
     GTAGGCCAGTGAGTGACCGGACCGCCGGTCAGGACCGTTGCATCACCGGTTTCGCGCCCA

     GATTACTGACTTGCAGGTCCAGGCCGGTGAGTTCCAGCCAGACGAAATTCTGGTTGCTCG
7261 ---------+---------+---------+---------+---------+---------+ 7320
     CTAATGACTGAACGTCCAGGTCCGGCCACTCAAGGTCGGTCTGCTTTAAGACCAACGAGC

     TTTTACCGATGCCAGTTGGACGCCGTTATTTGCCCTGGCGGCCGGTATTGTCACGGATAT
7321 ---------+---------+---------+---------+---------+---------+ 7380
     AAAATGGCTACGGTCAACCTGCGGCAATAAACGGGACCGCCGGCCATAACAGTGCCTATA

     TGGTTCCACACTTTCACACAGTTGTATTGTCGCGCGTGAGTTTGGGATCCCCGCCGTGGT
7381 ---------+---------+---------+---------+---------+---------+ 7440
     ACCAAGGTGTGAAAGTGTGTCAACATAACAGCGCGCACTCAAACCCTAGGGGCGGCACCA

     GAACCTGAAAACCGCCACCCAGATTATCAATAGTGGCGACATGTTGATCCTTGATGGAGA
7441 ---------+---------+---------+---------+---------+---------+ 7500
     CTTGGACTTTTGGCGGTGGGTCTAATAGTTATCACCGCTGTACAACTAGGAACTACCTCT

     TAGCGGGACGGTCATTATCCAACACCAAGAGGAGCGCAACCATGACGGTTAACCGGGCTG
7501 ---------+---------+---------+---------+---------+---------+ 7560
     ATCGCCCTGCCAGTAATAGGTTGTGGTTCTCCTCGCGTTGGTACTGCCAATTGGCCCGAC

     TCACGATCTGATTTAAATCACTGTAATGAATGCAATACAACGATGGCGGTGCGAGTGGTC
7561 ---------+---------+---------+---------+---------+---------+ 7620
     AGTGCTAGACTAAATTTAGTGACATTACTTACGTTATGTTGCTACCGCCACGCTCACCAG
```

```
     ATAGCCACGGCAACGTCGTTACGGGCTGTGTCGTGCCAAATTTGTCTGATTGGCCGACGT
7621 ---------+---------+---------+---------+---------+---------+ 7680
     TATCGGTGCCGTTGCAGCAATGCCCGACACAGCACGGTTTAAACAGACTAACCGGCTGCA

     CATAACTATATTTTGGCTGCATAAGGATTCGGTCATGACAACCATGATAGGACAAACAAG
7681 ---------+---------+---------+---------+---------+---------+ 7740
     GTATTGATATAAAACCGACGTATTCCTAAGCCAGTACTGTTGGTACTATCCTGTTTGTTC

     ACAAGCAGGCTCCTCTTCTTATGAGCAGGCTTGGCAAGCAGAACAAGCCCCCTGTCCAGG
7741 ---------+---------+---------+---------+---------+---------+ 7800
     TGTTCGTCCGAGGAGAAGAATACTCGTCCGAACCGTTCGTCTTGTTCGGGGGACAGGTCC

     GATGGAGCCCGATACCCTCACTGTGGGTGTGGTGGTAGTGACGCGCAATCCCACATTTTT
7801 ---------+---------+---------+---------+---------+---------+ 7860
     CTACCTCGGGCTATGGGAGTGACACCCACACCACCATCACTGCGCGTTAGGGTGTAAAAA

     TCAAACCGGGCTCAGCGTGTTAAATGATATTCGCGACTATGTGTTCAATCGTGTGCATAT
7861 ---------+---------+---------+---------+---------+---------+ 7920
     AGTTTGGCCCGAGTCGCACAATTTACTATAAGCGCTGATACACAAGTTAGCACACGTATA

     TCAATCGGAACTTCCGCTGAAGTTATCTGAATTAGCGTCGGACCCGCTTTATTCCGAAGC
7921 ---------+---------+---------+---------+---------+---------+ 7980
     AGTTAGCCTTGAAGGCGACTTCAATAGACTTAATCGCAGCCTGGGCGAAATAAGGCTTCG

     CCGCGAAAAAGCGATTCATTTTCTAAAAAATCAGAGCAAGGCGCTCAATATTCAAGTCAT
7981 ---------+---------+---------+---------+---------+---------+ 8040
     GGCGCTTTTTCGCTAAGTAAAAGATTTTTTAGTCTCGTTCCGCGAGTTATAAGTTCAGTA

     CCAGTGTGCCAGTCTTGCCGAAGCGACGGGGAAAATTATTTATACCCATGCTCTGGAACA
8041 ---------+---------+---------+---------+---------+---------+ 8100
     GGTCACACGGTCAGAACGGCTTCGCTGCCCCTTTTAATAAATATGGGTACGAGACCTTGT

     GCAGCCTGAATTTCAGATGGGAATGTTGTTTTACGATCAAACCTCGTTGGGGAATGTTGA
8101 ---------+---------+---------+---------+---------+---------+ 8160
     CGTCGGACTTAAAGTCTACCCTTACAACAAAATGCTAGTTTGGAGCAACCCCTTACAACT

     CGACAGCATTGAGAAAATTGACCGGGATCTGGACGCGTTTTATAGCGCCATGCAGAGAGG
8161 ---------+---------+---------+---------+---------+---------+ 8220
     GCTGTCGTAACTCTTTTAACTGGCCCTAGACCTGCGCAAAATATCGCGGTACGTCTCTCC

     TGGGATCCCCGCTTTTTATACCACTTTCTCAACAGTAACCTTTATTCGTGATGTCCGTTC
8221 ---------+---------+---------+---------+---------+---------+ 8280
     ACCCTAGGGGCGAAAAATATGGTGAAAGAGTTGTCATTGGAAATAAGCACTACAGGCAAG

     GTCGTTTCGTTATCTACCCCAGCAGTATCGAGAAATTGTCCGCAGTGAAGACCCCGCCAT
8281 ---------+---------+---------+---------+---------+---------+ 8340
     CAGCAAAGCAATAGATGGGGTCGTCATAGCTCTTTAACAGGCGTCACTTCTGGGGCGGTA

     CTTCCAAACCGAACTGCTGTGTTTGTGGATGGACTTTTTTGAGATGAATTATACCAACCG
8341 ---------+---------+---------+---------+---------+---------+ 8400
     GAAGGTTTGGCTTGACGACACAAACACCTACCTGAAAAAACTCTACTTAATATGGTTGGC

     CCGTGTCAAACCGATTGGGGCGTTGGCACTTCATAACACCCTGGCCGAGCAACTGATTCA
8401 ---------+---------+---------+---------+---------+---------+ 8460
     GGCACAGTTTGGCTAACCCCGCAACCGTGAAGTATTGTGGGACCGGCTCGTTGACTAAGT

     GTTCTTTGAGCGCACTGCCGCCAGCCGTTGGTTGGTGTCTTACTATACCGGTTCGATCAT
8461 ---------+---------+---------+---------+---------+---------+ 8520
     CAAGAAACTCGCGTGACGGCGGTCGGCAACCAACCACAGAATGATATGGCCAAGCTAGTA

     CTCTAATCTGATCGGTTATCTGGACAGACATGCTGAAGCGCATGGTGCGTTAGTTCTGCG
8521 ---------+---------+---------+---------+---------+---------+ 8580
     GAGATTAGACTAGCCAATAGACCTGTCTGTACGACTTCGCGTACCACGCAATCAAGACGC
```

```
      TGGCCCTAACGAACATGCCATTGCCTGTGGCGCCATGGCGAACTGGCAGCTTTATCGCAT
8581  ---------+---------+---------+---------+---------+---------+  8640
      ACCGGGATTGCTTGTACGGTAACGGACACCGCGGTACCGCTTGACCGTCGAAATAGCGTA

      GCCTTTTCTGGGGGTAGTGACCTCCGGCATGATGGATGAATTTAAAGGCACGCTGATCAA
8641  ---------+---------+---------+---------+---------+---------+  8700
      CGGAAAAGACCCCCATCACTGGAGGCCGTACTACCTACTTAAATTTCCGTGCGACTAGTT

      TCTGAAAGAGACGGCAGCACAGGGCATCATTGTTGCCGCAGAGAATCGTAATAATCAGTG
8701  ---------+---------+---------+---------+---------+---------+  8760
      AGACTTTCTCTGCCGTCGTGTCCCGTAGTAACAACGGCGTCTCTTAGCATTATTAGTCAC

      GTATAGCTTCCAGGGCACACAGACGCCGACTGAAGATATGCGCGATGTCCTGGCCGCCAA
8761  ---------+---------+---------+---------+---------+---------+  8820
      CATATCGAAGGTCCCGTGTGTCTGCGGCTGACTTCTATACGCGCTACAGGACCGGCGGTT

      GCGCATTCCGTATGTCTACATTGATGATGTTGACGGGATTGCTGATGGCCTGGCTGAAGT
8821  ---------+---------+---------+---------+---------+---------+  8880
      CGCGTAAGGCATACAGATGTAACTACTACAACTGCCCTAACGACTACCGGACCGACTTCA

      GTTTCGGCTCTACCATCAGGCCCAGGGGCCGGTAGTTATTTTAGCGACGCAAAATGTACT
8881  ---------+---------+---------+---------+---------+---------+  8940
      CAAAGCCGAGATGGTAGTCCGGGTCCCCGGCCATCAATAAAATCGCTGCGTTTTACATGA

      GGAATCGACTCTCTCGTTGGAGCCGGTGCCGGGTGACTTGCCGCCGGTTTCTGGTTTGCC
8941  ---------+---------+---------+---------+---------+---------+  9000
      CCTTAGCTGAGAGAGCAACCTCGGCCACGGCCCACTGAACGGCGGCCAAAGACCAAACGG

      AGCCTATGACTGCCCACCGATAAGCGACAGTTTTGAGCAAGCCATGGCGCTGATTAATGA
9001  ---------+---------+---------+---------+---------+---------+  9060
      TCGGATACTGACGGGTGGCTATTCGCTGTCAAAACTCGTTCGGTACCGCGACTAATTACT

      AGGGCCGGAAAAACTGGTTTGGCAGTTAGGACCGGTCAGTGACGATGAGTATGCCCTGGT
9061  ---------+---------+---------+---------+---------+---------+  9120
      TCCCGGCCTTTTTGACCAAACCGTCAATCCTGGCCAGTCACTGCTACTCATACGGGACCA

      TCATGAAATTGCCGATGCTGCCGGCCTGGCGCTGGTGGATTCGTTAGCGCATCCGGGTTC
9121  ---------+---------+---------+---------+---------+---------+  9180
      AGTACTTTAACGGCTACGACGGCCGGACCGCGACCACCTAAGCAATCGCGTAGGCCCAAG

      GGCACCGAAATATTATCAGGGCAAGCGCAATCCCCACTATTTGGGAACCCTGGCCATTTA
9181  ---------+---------+---------+---------+---------+---------+  9240
      CCGTGGCTTTATAATAGTCCCGTTCGCGTTAGGGGTGATAAACCCTTGGGACCGGTAAAT

      TGGTTATAGCCCCCGGGTTTACAACTTCCTGCATACCAACGACAAACTCAATCCAATGAG
9241  ---------+---------+---------+---------+---------+---------+  9300
      ACCAATATCGGGGGCCCAAATGTTGAAGGACGTATGGTTGCTGTTTGAGTTAGGTTACTC

      TGATCAAAGTGTGTTTATGATCAAAAGCCGTGTGGCGCAGATCACCACGCCGTTCTCTGA
9301  ---------+---------+---------+---------+---------+---------+  9360
      ACTAGTTTCACACAAATACTAGTTTTCGGCACACCGCGTCTAGTGGTGCGGCAAGAGACT

      TGGTCGGCTCGAGCGCAAGGTGCATTTGGTACAACTCACGCACGATGAACGACACTTATC
9361  ---------+---------+---------+---------+---------+---------+  9420
      ACCAGCCGAGCTCGCGTTCCACGTAAACCATGTTGAGTGCGTGCTACTTGCTGTGAATAG

      GCCGTACGCCGATTTGAAACTGCATATGGATTGCCTGACTTTCCTCCGGGCGGTGAAAGC
9421  ---------+---------+---------+---------+---------+---------+  9480
      CGGCATGCGGCTAAACTTTGACGTATACCTAACGGACTGAAAGGAGGCCCGCCACTTTCG

      CAATCTGCATGTCGATGCGGCACTACGGGAGAAACGCAAAGCGCTCATCGCTGCCTATCT
9481  ---------+---------+---------+---------+---------+---------+  9540
      GTTAGACGTACAGCTACGCCGTGATGCCCTCTTTGCGTTTCGCGAGTAGCGACGGATAGA
```

```
      GGATTCCCCTTCAGATGTTGTTAGCCAGTTACCCAGCCTGCCGATGTCGGCGAACTACTT
9541  ---------+---------+---------+---------+---------+---------+ 9600
      CCTAAGGGGAAGTCTACAACAATCGGTCAATGGGTCGGACGGCTACAGCCGCTTGATGAA

      TTTTTGCCAATTGAATCGGGTTATCGAGAATTTAATCAAGACGGAAAATTTCGATTTTAC
9601  ---------+---------+---------+---------+---------+---------+ 9660
      AAAAACGGTTAACTTAGCCCAATAGCTCTTAAATTAGTTCTGCCTTTTAAAGCTAAAATG

      CGGGGTTTACGATGTGGGGCGCTGTGGTATTTCAGCGGTACGCAATGTTGCCAAGACGCG
9661  ---------+---------+---------+---------+---------+---------+ 9720
      GCCCCAAATGCTACACCCCGCGACACCATAAAGTCGCCATGCGTTACAACGGTTCTGCGC

      GCGCGGTTTCTCGGGTTGGTATGGTCGGGCCTTGATGGGCGATGCATTACTGGCCACCAG
9721  ---------+---------+---------+---------+---------+---------+ 9780
      CGCGCCAAAGAGCCCAACCATACCAGCCCGGAACTACCCGCTACGTAATGACCGGTGGTC

      CTATCTGGCGCATACCAGTCCCACCCATGTGGTGGCGTTTATCGGCGACGGGGCCAAAGG
9781  ---------+---------+---------+---------+---------+---------+ 9840
      GATAGACCGCGTATGGTCAGGGTGGGTACACCACCGCAAATAGCCGCTGCCCCGGTTTCC

      GATTGTGCCGGATATTTTACCCGCCTTTATCGACAACATTCTCACCCATCCGCAGTTACT
9841  ---------+---------+---------+---------+---------+---------+ 9900
      CTAACACGGCCTATAAAATGGGCGGAAATAGCTGTTGTAAGAGTGGGTAGGCGTCAATGA

      CAATAAAAGCATCACCATTTTCTATTTTTGCAATGGCGGCTTGTCGGTCATCAACACCTA
9901  ---------+---------+---------+---------+---------+---------+ 9960
      GTTATTTTCGTAGTGGTAAAAGATAAAAACGTTACCGCCGAACAGCCAGTAGTTGTGGAT

      TCAGGAACGTATTTTGTTTAACCGCACATCGCGGCAGATGCGCCTGGTGAACGTTGATCA
9961  ---------+---------+---------+---------+---------+---------+ 10020
      AGTCCTTGCATAAAACAAATTGGCGTGTAGCGCCGTCTACGCGGACCACTTGCAACTAGT

      ACCCGCGTTTGAACAGACGGTGGATGATTTTCATATTCAGGGTAAAACACTCACTCATTT
10021 ---------+---------+---------+---------+---------+---------+ 10080
      TGGGCGCAAACTTGTCTGCCACCTACTAAAAGTATAAGTCCCATTTTGTGAGTGAGTAAA

      TGATGAGGACACCATTCGTCATGCATTGATGACCCCCAAACGACTGAACCTGTTTTCAGT
10081 ---------+---------+---------+---------+---------+---------+ 10140
      ACTACTCCTGTGGTAAGCAGTACGTAACTACTGGGGGTTTGCTGACTTGGACAAAAGTCA

      GGTACTGGGGCATAACAATGAAGGGGATGGCATTTCGTTAGCCACAGCCAAAGGCTGGCA
10141 ---------+---------+---------+---------+---------+---------+ 10200
      CCATGACCCCGTATTGTTACTTCCCCTACCGTAAAGCAATCGGTGTCGGTTTCCGACCGT

      ACGCGACCCCAGCGACCGGGAAGCGCTGCAAGAGCGTAAAGACTGGGCGGCTCGCCAGCC
10201 ---------+---------+---------+---------+---------+---------+ 10260
      TGCGCTGGGGTCGCTGGCCCTTCGCGACGTTCTCGCATTTCTGACCCGCCGAGCGGTCGG

      CGAATCGACAAGCACTTCATTCGACCAAGGTCAAAACAAGGAAGCTATATCATGAAGTTT
10261 ---------+---------+---------+---------+---------+---------+ 10320
      GCTTAGCTGTTCGTGAAGTAAGCTGGTTCCAGTTTTGTTCCTTCGATATAGTACTTCAAA

      GGATTTATCGCTCATCCAACCTCGGTTGGTTTAAAACGCTATGTCAAAATGATTGATTTA
10321 ---------+---------+---------+---------+---------+---------+ 10380
      CCTAAATAGCGAGTAGGTTGGAGCCAACCAAATTTTGCGATACAGTTTTACTAACTAAAT

      TTACAGCGTAATTCGACGGAGTTGCATAGCGGGTACAAACGAGACCTTTGGCGCCGGGAA
10381 ---------+---------+---------+---------+---------+---------+ 10440
      AATGTCGCATTAAGCTGCCTCAACGTATCGCCCATGTTTGCTCTGGAAACCGCGGCCCTT

      AACCTGGTCCCTTTCATGAACTTTGCCAAAATTACCTCAGCGACCGGCGCTACCTGTGAG
10441 ---------+---------+---------+---------+---------+---------+ 10500
      TTGGACCAGGGAAAGTACTTGAAACGGTTTTAATGGAGTCGCTGGCCGCGATGGACACTC
```

```
      GGAGTGATCAAATACATGCCGCTGGTCGCGGATGAGATGCTGGCTGATGCTCGCGGCATT
10501 ---------+---------+---------+---------+---------+---------+ 10560
      CCTCACTAGTTTATGTACGGCGACCAGCGCCTACTCTACGACCGACTACGAGCGCCGTAA

      GCCAATCGTGTGGTGTCGGGTATCGAAGAACTGGTCGAGGATGGCGCTGAACTGGTCGGT
10561 ---------+---------+---------+---------+---------+---------+ 10620
      CGGTTAGCACACCACAGCCCATAGCTTCTTGACCAGCTCCTACCGCGACTTGACCAGCCA

      TTGGGAGGATTCACCTCCATTGTCGGGCGGCGTGGTGAAGCCACCGCCGAGAAATCACCG
10621 ---------+---------+---------+---------+---------+---------+ 10680
      AACCCTCCTAAGTGGAGGTAACAGCCCGCCGCACCACTTCGGTGGCGGCTCTTTAGTGGC

      GTTCCGGTGACCTCCGGTAATTCGTTAACCACCTATGCGGGTTACAAAGCGCTGATGCAG
10681 ---------+---------+---------+---------+---------+---------+ 10740
      CAAGGCCACTGGAGGCCATTAAGCAATTGGTGGATACGCCCAATGTTTCGCGACTACGTC

      ATTCAGTCCTGGCTGGATATTCAGCCAGAGCAGGAGCCGGTTGCGATTGTCGGCTATCCG
10741 ---------+---------+---------+---------+---------+---------+ 10800
      TAAGTCAGGACCGACCTATAAGTCGGTCTCGTCCTCGGCCAACGCTAACAGCCGATAGGC

      GGCTCCATCTGTTTGGCGCTGAGCCGGTTATTGCTCGCCCAGGGTTTTTCCCTGCATTTG
10801 ---------+---------+---------+---------+---------+---------+ 10860
      CCGAGGTAGACAAACCGCGACTCGGCCAATAACGAGCGGGTCCCAAAAAGGGACGTAAAC

      CTGCATCGGGCCGGCCATAAAGATGAAGACGAATTGCTCAGCCATTTGCCGGAGCAATAT
10861 ---------+---------+---------+---------+---------+---------+ 10920
      GACGTAGCCCGGCCGGTATTTCTACTTCTGCTTAACGAGTCGGTAAACGGCCTCGTTATA

      CGCTCACGCGTCACGTTGACCAGTGATCCCGAGGATTTATACCCGCGTTGTAAACTGTTT
10921 ---------+---------+---------+---------+---------+---------+ 10980
      GCGAGTGCGCAGTGCAACTGGTCACTAGGGCTCCTAAATATGGGCGCAACATTTGACAAA

      GTCGCCGCAACATCCGCCGGAGGCGTGATTGACCCGTATAAACTGCAACCGGGTTCGGTC
10981 ---------+---------+---------+---------+---------+---------+ 11040
      CAGCGGCGTTGTAGGCGGCCTCCGCACTAACTGGGCATATTTGACGTTGGCCCAAGCCAG

      TTTATTGATGTGGCACTGCCCAGAGATATCAATTCTGACACGCGTCCCGATCGGGACGAC
11041 ---------+---------+---------+---------+---------+---------+ 11100
      AAATAACTACACCGTGACGGGTCTCTATAGTTAAGACTGTGCGCAGGGCTAGCCCTGCTG

      ATTTTGATTATCGACGGCGGCTGCGTTACGGCCACCGATGCGGTCAAACTGGGTGGTGAG
11101 ---------+---------+---------+---------+---------+---------+ 11160
      TAAAACTAATAGCTGCCGCCGACGCAATGCCGGTGGCTACGCCAGTTTGACCCACCACTC

      TCACTGAACGTCACCATTAAACAGCAATTGAATGGCTGTATGGCGGAAACCATTGTTTTG
11161 ---------+---------+---------+---------+---------+---------+ 11220
      AGTGACTTGCAGTGGTAATTTGTCGTTAACTTACCGACATACCGCCTTTGGTAACAAAAC

      GCACTGGAAAATCGCCGGGAGAATTTCTCGTTGGGCCGCTATCTGGCGCTGGATAACGTG
11221 ---------+---------+---------+---------+---------+---------+ 11280
      CGTGACCTTTTAGCGGCCCTCTTAAAGAGCAACCCGGCGATAGACCGCGACCTATTGCAC

      CTTGAAATCGGGGAACTTGCGGAAAAACATGGTTTCCTGGTCTATCCACTGGCCTCTTAT
11281 ---------+---------+---------+---------+---------+---------+ 11340
      GAACTTTAGCCCCTTGAACGCCTTTTTGTACCAAAGGACCAGATAGGTGACCGGAGAATA

      GGTGAGCGGATTGATCGTCAACGGGTGATCAATCTCAAGCGTTACTATCACCATGATATT
11341 ---------+---------+---------+---------+---------+---------+ 11400
      CCACTCGCCTAACTAGCAGTTGCCCACTAGTTAGAGTTCGCAATGATAGTGGTACTATAA

      TATTCCGATGAACCTGATACTGAACAACCGCCGGCGTCGCAGTTAGCCTTTATCGACGCG
11401 ---------+---------+---------+---------+---------+---------+ 11460
      ATAAGGCTACTTGGACTATGACTTGTTGGCGGCCGCAGCGTCAATCGGAAATAGCTGCGC
```

```
      ATTATTGCTCAGGATCCCGCCAGAGAAGACACTCTGGATCGCTATCATCAGTTCATTAAC
11461 ---------+---------+---------+---------+---------+---------+ 11520
      TAATAACGAGTCCTAGGGCGGTCTCTTCTGTGAGACCTAGCGATAGTAGTCAAGTAATTG

      CCGATGATGGTGGAATTTCTCAAACTGCAACACTGTGACAATGTATTCCGGCGGGCCTCG
11521 ---------+---------+---------+---------+---------+---------+ 11580
      GGCTACTACCACCTTAAAGAGTTTGACGTTGTGACACTGTTACATAAGGCCGCCCGGAGC

      GGTACTCAGCTGTTTACCGCTGACGGCGAAGCCTTTCTGGATATGGTGGCCGGTTATGGT
11581 ---------+---------+---------+---------+---------+---------+ 11640
      CCATGAGTCGACAAATGGCGACTGCCGCTTCGGAAAGACCTATACCACCGGCCAATACCA

      TGCATCAATCTGGGTCACAATCCGCAACCGATCATCGATGCTTTAAAAGCCTATCTCGAT
11641 ---------+---------+---------+---------+---------+---------+ 11700
      ACGTAGTTAGACCCAGTGTTAGGCGTTGGCTAGTAGCTACGAAATTTTCGGATAGAGCTA

      GCTCAGGGGCCCAATTTTATTCAGTATATTTCGATCCCCGAGCAAGCTGCCAAGTTGGCC
11701 ---------+---------+---------+---------+---------+---------+ 11760
      CGAGTCCCCGGGTTAAAATAAGTCATATAAAGCTAGGGGCTCGTTCGACGGTTCAACCGG

      GAAGTGCTCTGCCATTTCGCCCCTGGCAATATGGGGAGGGTGTTCTTCAGTAATTCAGGG
11761 ---------+---------+---------+---------+---------+---------+ 11820
      CTTCACGAGACGGTAAAGCGGGGACCGTTATACCCCTCCCACAAGAAGTCATTAAGTCCC

      ACGGAGGCGGTCGAGGCGGCAATGAAGCTGGCCAAAGCGTCGACGGGCAAAGCCGGCATT
11821 ---------+---------+---------+---------+---------+---------+ 11880
      TGCCTCCGCCAGCTCCGCCGTTACTTCGACCGGTTTCGCAGCTGCCCGTTTCGGCCGTAA

      GCGTATCTGAAAAATAGCTACCATGGCAAAACGCTGGGCGCGTTATCCATTACCGGACGA
11881 ---------+---------+---------+---------+---------+---------+ 11940
      CGCATAGACTTTTTATCGATGGTACCGTTTTGCGACCCGCGCAATAGGTAATGGCCTGCT

      GAAAAACACCGCCGTCATTTTAAGCCACTGCTGGCGTCGATGATTGAAGTACCGTTTGCT
11941 ---------+---------+---------+---------+---------+---------+ 12000
      CTTTTTGTGGCGGCAGTAAAATTCGGTGACGACCGCAGCTACTAACTTCATGGCAAACGA

      GATATTGAGGCACTGCGGCAGACCTTAAGCCGTGATGACATCGGCGCGTTGATGATCGAA
12001 ---------+---------+---------+---------+---------+---------+ 12060
      CTATAACTCCGTGACGCCGTCTGGAATTCGGCACTACTGTAGCCGCGCAACTACTAGCTT

      CCGATTCAGGGTGAAGGCGGGGTCCATGTTCCTCCGCCTGGATACCTGCGAACCGTTCAG
12061 ---------+---------+---------+---------+---------+---------+ 12120
      GGCTAAGTCCCACTTCCGCCCCAGGTACAAGGAGGCGGACCTATGGACGCTTGGCAAGTC

      GAGATCTGCCGTCAAACCGATACCTTACTGATGGTTGACGAAGTGCAAACCGGGTTGGGG
12121 ---------+---------+---------+---------+---------+---------+ 12180
      CTCTAGACGGCAGTTTGGCTATGGAATGACTACCAACTGCTTCACGTTTGGCCCAACCCC

      CGCACCGGGAAGCTCTTTGCCTGTGAGTGGGAAGGGATCGAGCCGGATGTACTGATGCTA
12181 ---------+---------+---------+---------+---------+---------+ 12240
      GCGTGGCCCTTCGAGAAACGGACACTCACCCTTCCCTAGCTCGGCCTACATGACTACGAT

      TCGAAATCACTGTCCGGCGGGGTAATGCCTATTGGCGCCACGCTATGTCGGGCCATATTT
12241 ---------+---------+---------+---------+---------+---------+ 12300
      AGCTTTAGTGACAGGCCGCCCCATTACGGATAACCGCGGTGCGATACAGCCCGGTATAAA

      GGCAACGGGCCGTATGGCACCGCAGACCGCTTTTTGATGCACAGCTCGACCTTTGGCGGC
12301 ---------+---------+---------+---------+---------+---------+ 12360
      CCGTTGCCCGGCATACCGTGGCGTCTGGCGAAAAACTACGTGTCGAGCTGGAAACCGCCG

      GGGAATATCGCGGCGGTAGTGGCGCTCAGTGCGTTAAGAGAAATTCTGGCTCAGGATCTG
12361 ---------+---------+---------+---------+---------+---------+ 12420
      CCCTTATAGCGCCGCCATCACCGCGAGTCACGCAATTCTCTTTAAGACCGAGTCCTAGAC
```

```
      GTCGGGAACGCCGAGCGACTAGGCACGTACTTTAAGCAGGCGCTTACCGATGTCGCTGCC
12421 --------+---------+---------+---------+---------+---------+ 12480
      CAGCCCTTGCGGCTCGCTGATCCGTGCATGAAATTCGTCCGCGAATGGCTACAGCGACGG

      CGATACCCCTTTGTGGCGGAAATAGCCGGACGCGGCTTGATGTTAGGGATCCAGTTCGAT
12481 --------+---------+---------+---------+---------+---------+ 12540
      GCTATGGGGAAACACCGCCTTTATCGGCCTGCGCCGAACTACAATCCCTAGGTCAAGCTA

      CAAACCTTCGCTGGTGCCGTAGGTGCTTCGGCCCGGGAGTTTGCCACCCGACTGCCCGGT
12541 --------+---------+---------+---------+---------+---------+ 12600
      GTTTGGAAGCGACCACGGCATCCACGAAGCCGGGCCCTCAAACGGTGGGCTGACGGGCCA

      GACTGGCACACGACATGGAAATTCCTGCCTGATCCGGTACAAGCCCACTTAAAGGCGGCG
12601 --------+---------+---------+---------+---------+---------+ 12660
      CTGACCGTGTGCTGTACCTTTAAGGACGGACTAGGCCATGTTCGGGTGAATTTCCGCCGC

      ATGGAGCGTATGGAACAATCACTGGGTGAAATGTTCTGCATGAAATTTGTGACCAAGCTT
12661 --------+---------+---------+---------+---------+---------+ 12720
      TACCTCGCATACCTTGTTAGTGACCCACTTTACAAGACGTACTTTAAACACTGGTTCGAA

      TGTCAGGATCACAACATTCTGACCTTTATTACCGCCAACAGCTCAACCGTTATTCGAATT
12721 --------+---------+---------+---------+---------+---------+ 12780
      ACAGTCCTAGTGTTGTAAGACTGGAAATAATGGCGGTTGTCGAGTTGGCAATAAGCTTAA

      CAACCGCCACTGACCATCAGCAAGGCTGAGATCGATCGTTTTGTCAGTGCGTTTGCCACG
12781 --------+---------+---------+---------+---------+---------+ 12840
      GTTGGCGGTGACTGGTAGTCGTTCCGACTCTAGCTAGCAAAACAGTCACGCAAACGGTGC

      GTGTGCGATGAGCTATCAACATTTTTAGAGTAAGGACGCTGAAATGACATTAACCAAGCA
12841 --------+---------+---------+---------+---------+---------+ 12900
      CACACGCTACTCGATAGTTGTAAAAATCTCATTCCTGCGACTTTACTGTAATTGGTTCGT

      AGACGCAGTCAACCAGATGATGGGCTTTTTTCAGTCCAAAACGCTGATCACGGCGCTGTC
12901 --------+---------+---------+---------+---------+---------+ 12960
      TCTGCGTCAGTTGGTCTACTACCCGAAAAAAGTCAGGTTTTGCGACTAGTGCCGCGACAG

      ATTAAAGTTGTTTGATCACCTGCGTGATCAGGATCGCAACGCTAAACAGATGGCCGCATT
12961 --------+---------+---------+---------+---------+---------+ 13020
      TAATTTCAACAAACTAGTGGACGCACTAGTCCTAGCGTTGCGATTTGTCTACCGGCGTAA

      ACTCAACTGTCCTTTGCGTTCGAGCGAGCAGTTACTCATTGCCTTGCAGGCGATGGGATA
13021 --------+---------+---------+---------+---------+---------+ 13080
      TGAGTTGACAGGAAACGCAAGCTCGCTCGTCAATGAGTAACGGAACGTCCGCTACCCTAT

      TCTTGAAAAACAGGACGGTCTTTATCATTTGCCGCAGGAACACCGGGCGTTTTTGGTCAG
13081 --------+---------+---------+---------+---------+---------+ 13140
      AGAACTTTTTGTCCTGCCAGAAATAGTAAACGGCGTCCTTGTGGCCCGCAAAAACCAGTC

      CGATGAACCACAGTGGCTGGGATGGCTGGGCCGTCATATCGACACTTTTCTATACCCGCT
13141 --------+---------+---------+---------+---------+---------+ 13200
      GCTACTTGGTGTCACCGACCCTACCGACCCGGCAGTATAGCTGTGAAAAGATATGGGCGA

      ATGGGGAGAGTTGAAAGCGGCTGTTGAAAACGATACCCACCAGCGCCAGACCGTCTTCGG
13201 --------+---------+---------+---------+---------+---------+ 13260
      TACCCCTCTCAACTTTCGCCGACAACTTTTGCTATGGGTGGTCGCGGTCTGGCAGAAGCC

      CGATGATCGAAGCTGGTTCGATATTTTGTACCAAAACCCGGATGACGTGACTGATTTCCA
13261 --------+---------+---------+---------+---------+---------+ 13320
      GCTACTAGCTTCGACCAAGCTATAAAACATGGTTTTGGGCCTACTGCACTGACTAAAGGT

      GGAGTTTCTGGGGAAGTTTGCCGCCCCCTTCATCGACGGGTTTATTCAGGATTATGACTT
13321 --------+---------+---------+---------+---------+---------+ 13380
      CCTCAAAGACCCCTTCAAACGGCGGGGGAAGTAGCTGCCCAAATAAGTCCTAATACTGAA
```

```
      CTCGCAGCATCAGGCATTTTTAGATATTGGCAGTGGAATTGGGAGCTTGCCTATCGCAGT
13381 ---------+---------+---------+---------+---------+---------+ 13440
      GAGCGTCGTAGTCCGTAAAAATCTATAACCGTCACCTTAACCCTCGAACGGATAGCGTCA

      AGCCAATGCCTATTCGGGCGTGAATCTGGCGATTTGCGAGCTGCCCCAAACGTCGACGTT
13441 ---------+---------+---------+---------+---------+---------+ 13500
      TCGGTTACGGATAAGCCCGCACTTAGACCGCTAAACGCTCGACGGGGTTTGCAGCTGCAA

      TTTGCGTGACAAGTTGGTTCAGCAAGGTTATGGGCAGCGGATCCAGGTGCTTGAGGGGGA
13501 ---------+---------+---------+---------+---------+---------+ 13560
      AAACGCACTGTTCAACCAAGTCGTTCCAATACCCGTCGCCTAGGTCCACGAACTCCCCCT

      TGTGATTAGCGGTGATTTGCCGATTGGCGATTATGACCTGATCCACCTGGGATGGATGCT
13561 ---------+---------+---------+---------+---------+---------+ 13620
      ACACTAATCGCCACTAAACGGCTAACCGCTAATACTGGACTAGGTGGACCCTACCTACGA

      GCATGACTACGCTCCTGAAACGCAATTGATCATCTTGAAAAATATCTATGATGCGATGCC
13621 ---------+---------+---------+---------+---------+---------+ 13680
      CGTACTGATGCGAGGACTTTGCGTTAACTAGTAGAACTTTTTATAGATACTACGCTACGG

      GGTCGGTGGTCGGTTTATCGCCTCTGAAACGCCGCTGAATGCGGATAAGTCAGGCCCTGA
13681 ---------+---------+---------+---------+---------+---------+ 13740
      CCAGCCACCAGCCAAATAGCGGAGACTTTGCGGCGACTTACGCCTATTCAGTCCGGGACT

      GTTTACCGCACTGTTATCGCTGAACATGCTGGTATCCACCGATGGCGGTATCGAAAGCAG
13741 ---------+---------+---------+---------+---------+---------+ 13800
      CAAATGGCGTGACAATAGCGACTTGTACGACCATAGGTGGCTACCGCCATAGCTTTCGTC

      TCCTCAGGAGTATTTGTCCCGGTTTCATCAGGCAGGATTCAGCAATGCCCGGATCATGGA
13801 ---------+---------+---------+---------+---------+---------+ 13860
      AGGAGTCCTCATAAACAGGGCCAAAGTAGTCCGTCCTAAGTCGTTACGGGCCTAGTACCT

      TATATCCGGTCCTCGTACACTGATAGTCGGCGAGAAAACAACGCATAACAACGGGAGTTC
13861 ---------+---------+---------+---------+---------+---------+ 13920
      ATATAGGCCAGGAGCATGTGACTATCAGCCGCTCTTTTGTTGCGTATTGTTGCCCTCAAG

      ACAATGTTAGAAAGTAAATTGATAAACCATATTGCTACTCAGTTTTTGGATGGTGAAAAG
13921 ---------+---------+---------+---------+---------+---------+ 13980
      TGTTACAATCTTTCATTTAACTATTTGGTATAACGATGAGTCAAAAACCTACCACTTTTC

      GATGGCCTGGATAGCCAAACACCCTTGTTTGAGCTGAATATAGTTGACTCGGCTGCAATT
13981 ---------+---------+---------+---------+---------+---------+ 14040
      CTACCGGACCTATCGGTTTGTGGGAACAAACTCGACTTATATCAACTGAGCCGACGTTAA

      TTTGATCTGGTGGATTTTTTAAGGCAAGAGAGCAAGGTCTCGATTGGCATGCAAGAGATT
14041 ---------+---------+---------+---------+---------+---------+ 14100
      AAACTAGACCACCTAAAAAATTCCGTTCTCTCGTTCCAGAGCTAACCGTACGTTCTCTAA

      CACCCGGCAAATTTCGCCACCGTGCAGAGTATGGTTGCGCTGGTGCAACGGCTGAAGGCG
14101 ---------+---------+---------+---------+---------+---------+ 14160
      GTGGGCCGTTTAAAGCGGTGGCACGTCTCATACCAACGCGACCACGTTGCCGACTTCCGC

      CATCCGGAGCAGGGAGGTGCGGCATGAACGATGTAACGACCGAGACCTATGAAACCCTGA
14161 ---------+---------+---------+---------+---------+---------+ 14220
      GTAGGCCTCGTCCCTCCACGCCGTACTTGCTACATTGCTGGCTCTGGATACTTTGGGACT

      AACAATCTGTATTGCACACCTTCGCGCAATTGACCGGATACAACGTGTCAGAGCTATCGT
14221 ---------+---------+---------+---------+---------+---------+ 14280
      TTGTTAGACATAACGTGTGGAAGCGCGTTAACTGGCCTATGTTGCACAGTCTCGATAGCA

      TAACCAGCCATTTGGAAAATGATCTGGGGGTGGATTCCATCGCGTTGGCAGAGATTGCCG
14281 ---------+---------+---------+---------+---------+---------+ 14340
      ATTGGTCGGTAAACCTTTTACTAGACCCCCACCTAAGGTAGCGCAACCGTCTCTAACGGC
```

APPENDIX IV

THE COMPLETE PRODIGIOSIN BIOSYNTHETIC CLUSTER DNA SEQUENCE AND PUTATIVE PROTEINS ENCODED BY THE 5' END OF THE CLUSTER

```
      TCTCGCTGAGCAGACAGTTTCAGCTTAATACGCCATTGTTGATTCAGGATATAAACACCA
14341 --------+---------+---------+---------+---------+---------+ 14400
      AGAGCGACTCGTCTGTCAAAGTCGAATTATGCGGTAACAACTAAGTCCTATATTTGTGGT

      TCAAGGACGCGCTTGACGGCATTTTGCAACGCGAGTTTCAATTGTCGGAAAAGGTCGAGC
14401 --------+---------+---------+---------+---------+---------+ 14460
      AGTTCCTGCGCGAACTGCCGTAAAACGTTGCGCTCAAAGTTAACAGCCTTTTCCAGCTCG

      CGGCAGCAATTGCTCTCTCCGGGGACGCTGACCTTTGGCTTGGTAATCTTGTCCGTCAGA
14461 --------+---------+---------+---------+---------+---------+ 14520
      GCCGTCGTTAACGAGAGAGGCCCCTGCGACTGGAAACCGAACCATTAGAACAGGCAGTCT

      TCTTTGCCAGCCATAGTGGCTATGATGTTAATGCGCTGGCGCTGGATGCTGAAATCGAAA
14521 --------+---------+---------+---------+---------+---------+ 14580
      AGAAACGGTCGGTATCACCGATACTACAATTACGCGACCGCGACCTACGACTTTAGCTTT

      GTGATCTGGGGATCGACTCGGTGTCGGTGGCGTCCGCTCAGGGTGAACTATTCAATACGC
14581 --------+---------+---------+---------+---------+---------+ 14640
      CACTAGACCCCTAGCTGAGCCACAGCCACCGCAGGCGAGTCCCACTTGATAAGTTATGCG

      TGCAATTAAATAGCGAAACCATCATCGCGAACTGCAACACGCTGTCAGCGTTGAAGCAGT
14641 --------+---------+---------+---------+---------+---------+ 14700
      ACGTTAATTTATCGCTTTGGTAGTAGCGCTTGACGTTGTGCGACAGTCGCAACTTCGTCA

      GTTTAGCGGCGCGTCTGGTGCAAGAAAAAGGTCAGGATTGGTTTGAACAACGTGGGAGAG
14701 --------+---------+---------+---------+---------+---------+ 14760
      CAAATCGCCGCGCAGACCACGTTCTTTTTCCAGTCCTAACCAAACTTGTTGCACCCTCTC

      GGCAATCCGACTCCGCCATTGACCATGATGCCGACACGACGGCGGAGGTGACTCCGCCAA
14761 --------+---------+---------+---------+---------+---------+ 14820
      CCGTTAGGCTGAGGCGGTAACTGGTACTACGGCTGTGCTGCCGCCTCCACTGAGGCGGTT

      CAGCGACACCCGTCGCGATCAATGCCGAGATCGGCGATCCACGCACCATGCGTGATTTTG
14821 --------+---------+---------+---------+---------+---------+ 14880
      GTCGCTGTGGGCAGCGCTAGTTACGGCTCTAGCCGCTAGGTGCGTGGTACGCACTAAAAC

      TGGGAATAGAACATCCGGATATTTTCCACAAGGCGCGGGAATTCCACCTCTTTTATCAGG
14881 --------+---------+---------+---------+---------+---------+ 14940
      ACCCTTATCTTGTAGGCCTATAAAAGGTGTTCCGCGCCCTTAAGGTGGAGAAAATAGTCC

      ATAAAAAGAAACGGCAGCTCTATTTCTATGGCATGCCGCTGGAGACGCCGTGTAAAAACC
14941 --------+---------+---------+---------+---------+---------+ 15000
      TATTTTTCTTTGCCGTCGAGATAAAGATACCGTACGGCGACCTCTGCGGCACATTTTTGG

      GGGCCGTTATGTTTGATGAAGCAACCGGTCAACACCGTGAGTTTTTAATGTTTGGCTCCA
15001 --------+---------+---------+---------+---------+---------+ 15060
      CCCGGCAATACAAACTACTTCGTTGGCCAGTTGTGGCACTCAAAAATTACAAACCGAGGT

      ACAGCTACCTGGGGTTGTCGAATCATCCGGAAATCATTCATGCCATCCAGGATGCGGCCA
15061 --------+---------+---------+---------+---------+---------+ 15120
      TGTCGATGGACCCCAACAGCTTAGTAGGCCTTTAGTAAGTACGGTAGGTCCTACGCCGGT

      GCTTATATGGGGCCACCAATACCGGCTGTCGTATTATCGCTGGCAGTAATGTGCTGCATT
15121 --------+---------+---------+---------+---------+---------+ 15180
      CGAATATACCCCGGTGGTTATGGCCGACAGCATAATAGCGACCGTCATTACACGACGTAA

      TAGAGCTGGAGCGCAAACTGGCCAAACTTAAAGGTCGTGATGATTGCATTGTGTATCCTT
15181 --------+---------+---------+---------+---------+---------+ 15240
      ATCTCGACCTCGCGTTTGACCGGTTTGAATTTCCAGCACTACTAACGTAACACATAGGAA

      CCGGTTACTCAGCCAACCTGGGGTGCATCTCGGCGCTGACCAGCAGACACGATTTGGTGT
15241 --------+---------+---------+---------+---------+---------+ 15300
      GGCCAATGAGTCGGTTGGACCCCACGTAGAGCCGCGACTGGTCGTCTGTGCTAAACCACA
```

```
      TTACCGATGCCATCAATCACATGAGTATTCAGGATGGCTGTAAGCTGGCTGGCGCGCAGC
15301 --------+---------+---------+---------+---------+---------+ 15360
      AATGGCTACGGTAGTTAGTGTACTCATAAGTCCTACCGACATTCGACCGACCGCGCGTCG

      GCAAGATCTACAATCACTCCTTGACCAGTCTGGAAAAATCCCTGGCCAAATATGCTGATC
15361 --------+---------+---------+---------+---------+---------+ 15420
      CGTTCTAGATGTTAGTGAGGAACTGGTCAGACCTTTTTAGGGACCGGTTTATACGACTAG

      ATCCTGGCGGTAAACTGATTGTGACCGATGGGGTGTTCAGTATGCATGGCGACATTGTCG
15421 --------+---------+---------+---------+---------+---------+ 15480
      TAGGACCGCCATTTGACTAACACTGGCTACCCCACAAGTCATACGTACCGCTGTAACAGC

      ACCTGCCCAGGTTGATGAAGCTGGCCGAGCGTTACGGCGCACGCGTGTTGGTGGATGATG
15481 --------+---------+---------+---------+---------+---------+ 15540
      TGGACGGGTCCAACTACTTCGACCGGCTCGCAATGCCGCGTGCGCACAACCACCTACTAC

      CGCACTCCACTGGCGTTTTAGGGAAAACCGGGGCGGGGACATCGGAACACTTCAATATGA
15541 --------+---------+---------+---------+---------+---------+ 15600
      GCGTGAGGTGACCGCAAAATCCCTTTTGGCCCCGCCCCTGTAGCCTTGTGAAGTTATACT

      AAGGTCAGGTGGATCTGGAACTGGGAACCATGAGTAAAGCGCTGTCCGGTCTGGGAGGCT
15601 --------+---------+---------+---------+---------+---------+ 15660
      TTCCAGTCCACCTAGACCTTGACCCTTGGTACTCATTTCGCGACAGGCCAGACCCTCCGA

      ATGTCTGTGGTGATGGCGACGTGGTTGAGTATTTGCGCTTTTATTCCAACTCCTATGTCT
15661 --------+---------+---------+---------+---------+---------+ 15720
      TACAGACACCACTACCGCTGCACCAACTCATAAACGCGAAAATAAGGTTGAGGATACAGA

      TTGCCGCCACCATTCCGGCACCCGTGGCGGCGGGAGTGATCGCCTCAATTGATGTGATGC
15721 --------+---------+---------+---------+---------+---------+ 15780
      AACGGCGGTGGTAAGGCCGTGGGCACCGCCGCCCTCACTAGCGGAGTTAACTACACTACG

      TGCGTGAACCCGAGCGTTTAGCCAAACTGTGGGACAATATTTACTACTTCCGCACCCGAT
15781 --------+---------+---------+---------+---------+---------+ 15840
      ACGCACTTGGGCTCGCAAATCGGTTTGACACCCTGTTATAAATGATGAAGGCGTGGGCTA

      TACTCAATGCCGGTTTTGATCTTGAGAACTCTGATTCGGCCATTATTCCGATTGTGGTCG
15841 --------+---------+---------+---------+---------+---------+ 15900
      ATGAGTTACGGCCAAAACTAGAACTCTTGAGACTAAGCCGGTAATAAGGCTAACACCAGC

      GCGATGATGCCAAGACGCTCTTCTTTGGTCGAGCGGTTCGGGCGCGGGGAATGTTTTGCC
15901 --------+---------+---------+---------+---------+---------+ 15960
      CGCTACTACGGTTCTGCGAGAAGAAACCAGCTCGCCAAGCCCGCGCCCCTTACAAAACGG

      AAACGGTCGTCTTCCCTGGTGTCAGCGTAGGGGATGCGCGGTTACGCATTAGCATCACCT
15961 --------+---------+---------+---------+---------+---------+ 16020
      TTTGCCAGCAGAAGGGACCACAGTCGCATCCCCTACGCGCCAATGCGTAATCGTAGTGGA

      CTGAACACACTCGGGAAGATTTGGACGAAGCGTATGCGATCCTGGTGGCATCGGCTCTGG
16021 --------+---------+---------+---------+---------+---------+ 16080
      GACTTGTGTGAGCCCTTCTAAACCTGCTTCGCATACGCTAGGACCACCGTAGCCGAGACC

      AAGTCGGTGTGCCGGTTAATGCATCCGCTCATCAGGAAGAGAACGCGAGTGTCGCGGAGG
16081 --------+---------+---------+---------+---------+---------+ 16140
      TTCAGCCACACGGCCAATTACGTAGGCGAGTAGTCCTTCTCTTGCGCTCACAGCGCCTCC

      CTTAACCGATGACAATATCCACTCCCGTGATTATCGACAGTCTCATCAGGCACGCCCAGC
16141 --------+---------+---------+---------+---------+---------+ 16200
      GAATTGGCTACTGTTATAGGTGAGGGCACTAATAGCTGTCAGAGTAGTCCGTGCGGGTCG

      GAACGCCAGAACAGACAGCCTTGTTGTGTGGTGATCAACACTGGAATTATCGGCAATTGG
16201 --------+---------+---------+---------+---------+---------+ 16260
      CTTGCGGTCTTGTCTGTCGGAACAACACACCACTAGTTGTGACCTTAATAGCCGTTAACC
```

```
      TCACTCGTGCTCACGTTATGGCCTCGGCCCTGCGACAGGCCGGATTATCGGGTCAGGCCA
16261 --------+---------+---------+---------+---------+---------+ 16320
      AGTGAGCACGAGTGCAATACCGGAGCCGGGACGCTGTCCGGCCTAATAGCCCAGTCCGGT

      TTTTGCTTAATCTGCCGAAAAGTCTGGATGCGGTTGCGGCGATATATGCCACCTGGCTAA
16321 --------+---------+---------+---------+---------+---------+ 16380
      AAAACGAATTAGACGGCTTTTCAGACCTACGCCAACGCCGCTATATACGGTGGACCGATT

      GTGGTAACCACTACATTCCTATTGACTATAGTCAGCCGTCATCACGTATTGAACGCATTA
16381 --------+---------+---------+---------+---------+---------+ 16440
      CACCATTGGTGATGTAAGGATAACTGATATCAGTCGGCAGTAGTGCATAACTTGCGTAAT

      TTGCCGCTGCCGCCCCGGCGTTGATTATTGATACAGCGTGGTTAGCCACACTCGATAGCC
16441 --------+---------+---------+---------+---------+---------+ 16500
      AACGGCGACGGCGGGGCCGCAACTAATAACTATGTCGCACCAATCGGTGTGAGCTATCGG

      AACCCTCTTTCGATGCGGAGCAACCAGTTGGGCGTATGGTGTACCACAATCCCATCGCTG
16501 --------+---------+---------+---------+---------+---------+ 16560
      TTGGGAGAAAGCTACGCCTCGTTGGTCAACCCGCATACCACATGGTGTTAGGGTAGCGAC

      CCATTTTATATACCTCCGGCTCGACAGGCACCCCGAAAGGGGTGCAAATCAGCCACGAAA
16561 --------+---------+---------+---------+---------+---------+ 16620
      GGTAAAATATATGGAGGCCGAGCTGTCCGTGGGGCTTTCCCCACGTTTAGTCGGTGCTTT

      TGCTGGGCTTTTTTATCCAGTGGGCGGTGCGGGATACGCAACTGACAGCGCGGGATGTGC
16621 --------+---------+---------+---------+---------+---------+ 16680
      ACGACCCGAAAAAATAGGTCACCCGCCACGCCCTATGCGTTGACTGTCGCGCCCTACACG

      TGTCCAATCACGCCAGTTTTGCGTTTGATTTAAGTACCTTTGACCTGTTTGCCAGTGCTT
16681 --------+---------+---------+---------+---------+---------+ 16740
      ACAGGTTAGTGCGGTCAAAACGCAAACTAAATTCATGGAAACTGGACAAACGGTCACGAA

      ATGTCGGTGCTGCCACCTGGATTATTCGCGAAAGCGAACAGAAAGATTGCGCCGCGTTGG
16741 --------+---------+---------+---------+---------+---------+ 16800
      TACAGCCACGACGGTGGACCTAATAAGCGCTTTCGCTTGTCTTTCTAACGCGGCGCAACC

      CACAGGGGCTGCAAAGACATGCGGTGTCGGTGTGGTACAGCGTGCCTTCCATTTTAGCCA
16801 --------+---------+---------+---------+---------+---------+ 16860
      GTGTCCCCGACGTTTCTGTACGCCACAGCCACACCATGTCGCACGGAAGGTAAAATCGGT

      TGCTGGAAAAAAGTACACTGTTGAACCCGACGCTCGGCCAATCTCTGCGGCAAGTCATTT
16861 --------+---------+---------+---------+---------+---------+ 16920
      ACGACCTTTTTTCATGTGACAACTTGGGCTGCGAGCCGGTTAGAGACGCCGTTCAGTAAA

      TCGCTGGCGAGCCCTATCCGGTCACTGCGTTAAAACGGCTCTTGCCTTGTTTGCCTCAGC
16921 --------+---------+---------+---------+---------+---------+ 16980
      AGCGACCGCTCGGGATAGGCCAGTGACGCAATTTTGCCGAGAACGGAACAAACGGAGTCG

      CATGCCGGGTCAGTAACTGGTATGGCCCCACAGAAACCAATGTCTGCGTGGCCTATGCGA
16981 --------+---------+---------+---------+---------+---------+ 17040
      GTACGGCCCAGTCATTGACCATACCGGGGTGTCTTTGGTTACAGACGCACCGGATACGCT

      TTGATCGGGCGCGACTGGCTATGTTGAAACAGGTGCCGATTGGACTTCCCCTTGAAGGAC
17041 --------+---------+---------+---------+---------+---------+ 17100
      AACTAGCCCGCGCTGACCGATACAACTTTGTCCACGGCTAACCTGAAGGGGAACTTCCTG

      TGACGGCACAGCTCGAAGATGAAAATGGTGACCGACATCCGCTGACGGCACAACTGCGCT
17101 --------+---------+---------+---------+---------+---------+ 17160
      ACTGCCGTGTCGAGCTTCTACTTTTACCACTGGCTGTAGGCGACTGCCGTGTTGACGCGA

      TAAGCGGCGAGTTGCTGATCAGCGGGCCGTGTGTGACGCCGGGCTACAGCAATGTGGTCG
17161 --------+---------+---------+---------+---------+---------+ 17220
      ATTCGCCGCTCAACGACTAGTCGCCCGGCACACACTGCGGCCCGATGTCGTTACACCAGC
```

```
      TTCCCCGGCAGGCCGCTTTGCATCCACATCAATGCCATGCCACCGGGGACTGGGTAGAGA
17221 --------+---------+---------+---------+---------+---------+ 17280
      AAGGGGCCGTCCGGCGAAACGTAGGTGTAGTTACGGTACGGTGGCCCCTGACCCATCTCT

      TGACGCCAGAGGGTCTGGTATTCCGTGGCCGTATTGATGATATGGTCAAAATCAATGGTT
17281 --------+---------+---------+---------+---------+---------+ 17340
      ACTGCGGTCTCCCAGACCATAAGGCACCGGCATAACTACTATACCAGTTTTAGTTACCAA

      ATCGAGTGGAGTTAGGCGAGATTGAATCGGTACTTCATCAGCATCCGGCGATTGATCGGG
17341 --------+---------+---------+---------+---------+---------+ 17400
      TAGCTCACCTCAATCCGCTCTAACTTAGCCATGAAGTAGTCGTAGGCCGCTAACTAGCCC

      CGGCGCTATGTGTCGAACTTGGGGATCTCCGGCAGACGTTAATCATGGTGATCAGTCTAC
17401 --------+---------+---------+---------+---------+---------+ 17460
      GCCGCGATACACAGCTTGAACCCCTAGAGGCCGTCTGCAATTAGTACCACTAGTCAGATG

      AGACTGGCGCAGTGCCGCCGGGTCTGCTGGAACTAAAACAGTTCCTGCAACAGAAACTTC
17461 --------+---------+---------+---------+---------+---------+ 17520
      TCTGACCGCGTCACGGCGGCCCAGACGACCTTGATTTTGTCAAGGACGTTGTCTTTGAAG

      CCTCTTATATGATCCCCAATAAGCTGGTGATCACCGAGAGTTTGCCGGTTAACGCCAATG
17521 --------+---------+---------+---------+---------+---------+ 17580
      GGAGAATATACTAGGGGTTATTCGACCACTAGTGGCTCTCAAACGGCCAATTGCGGTTAC

      GCAAGGTTGATCGTAAGCAGTTAGCCGGGGTGGTTGCGGTATGAGTAATGATAAACACAT
17581 --------+---------+---------+---------+---------+---------+ 17640
      CGTTCCAACTAGCATTCGTCAATCGGCCCCACCAACGCCATACTCATTACTATTTGTGTA

      CGCACCACTGGCGGTGGTCAGCATGGGGTGCGTGTTGCCCGGTGTTGACCATTTTCGGGC
17641 --------+---------+---------+---------+---------+---------+ 17700
      GCGTGGTGACCGCCACCAGTCGTACCCCACGCACAACGGGCCACAACTGGTAAAAGCCCG

      GCTAGACACTATTGCGGATTGGGAAACGGTGTTTCAGTCGGCGTCACCGCTGGCGTGGAG
17701 --------+---------+---------+---------+---------+---------+ 17760
      CGATCTGTGATAACGCCTAACCCTTTGCCACAAAGTCAGCCGCAGTGGCGACCGCACCTC

      CGAGACATCCCGACCAATTCAAGGTCGGCAGATGGATGATTCCGGGTTTGATTTTAAAAA
17761 --------+---------+---------+---------+---------+---------+ 17820
      GCTCTGTAGGGCTGGTTAAGTTCCAGCCGTCTACCTACTAAGGCCCAAACTAAAATTTTT

      GTTTTCTATTCCGCCTTTATTCAGGAAAGCCGTTAGCCGGGAAACCCGTTTGGCACTGCG
17821 --------+---------+---------+---------+---------+---------+ 17880
      CAAAAGATAAGGCGGAAATAAGTCCTTTCGGCAATCGGCCCTTTGGGCAAACCGTGACGC

      GGCTGCCGAAGATGCTCTGGCCGGGCTGGTGTTGCCGGAATCACTGCGTGACTGCTGCGA
17881 --------+---------+---------+---------+---------+---------+ 17940
      CCGACGGCTTCTACGAGACCGGCCCGACCACAACGGCCTTAGTGACGCACTGACGACGCT

      TCAGTTTTGTGCCATTCACCTGGGCAGTGATGCCGCCTACCGCAATGCGACGAAAGTGGG
17941 --------+---------+---------+---------+---------+---------+ 18000
      AGTCAAAACACGGTAAGTGGACCCGTCACTACGGCGGATGGCGTTACGCTGCTTTCACCC

      CGCGTTACGGGCATTAGCCGAAAAACTTCAGGCGCAAGGTTGCCCGGCGGCGGAGGTGAG
18001 --------+---------+---------+---------+---------+---------+ 18060
      GCGCAATGCCCGTAATCGGCTTTTTGAAGTCCGCGTTCCAACGGGCCGCCGCCTCCACTC

      GCGTCGGCTTGACGACTATAAACAACCGCTGGCGGAGTCGCTTGGGTGTTCATCTCATGA
18061 --------+---------+---------+---------+---------+---------+ 18120
      CGCAGCCGAACTGCTGATATTTGTTGGCGACCGCCTCAGCGAACCCACAAGTAGAGTACT

      TCGGGTGGGGGAAATGGCCTCCAGCATTCCTGCGCGAATCGCTCACTTTGCTCATACGCG
18121 --------+---------+---------+---------+---------+---------+ 18180
      AGCCCACCCCCTTTACCGGAGGTCGTAAGGACGCGCTTAGCGAGTGAAACGAGTATGCGC
```

```
      CGGTAAATGTCAGACCCTCGACGGTGCTGACAAGGGCGGGCTACGTTTATTGCAACTGGC
18181 ---------+---------+---------+---------+---------+---------+ 18240
      GCCATTTACAGTCTGGGAGCTGCCACGACTGTTCCCGCCCGATGCAAATAACGTTGACCG

      TCAGGACTGCTTTCGCTATCATGATAGCCAGATGGCGGTGTTGACTTCTGTTCAATGCTT
18241 ---------+---------+---------+---------+---------+---------+ 18300
      AGTCCTGACGAAAGCGATAGTACTATCGGTCTACCGCCACAACTGAAGACAAGTTACGAA

      TCACCATCGTCCCCAGGCGTACATGCTGCTCGAACAAGGGGTATCGCAGGATGCCTGTTG
18301 ---------+---------+---------+---------+---------+---------+ 18360
      AGTGGTAGCAGGGGTCCGCATGTACGACGAGCTTGTTCCCCATAGCGTCCTACGGACAAC

      GCTGGAAGGGGCGATCAGCCTGGTGGTGTGTCCCCTCGCGGTGGCGCATGAGCAAGGGTG
18361 ---------+---------+---------+---------+---------+---------+ 18420
      CGACCTTCCCCGCTAGTCGGACCACCACACAGGGGAGCGCCACCGCGTACTCGTTCCCAC

      GCCAGTGCTGACCCAACTCGGGGATATTGTGACCACCCATGATGGGTCACCGCAACCTGA
18421 ---------+---------+---------+---------+---------+---------+ 18480
      CGGTCACGACTGGGTTGAGCCCCTATAACACTGGTGGGTACTACCCAGTGGCGTTGGACT

      GGCAGACCATCCGGCCGCGTTATACTTTGCCGGAGCCAACCAGGTTTTTTGCCAGATCGT
18481 ---------+---------+---------+---------+---------+---------+ 18540
      CCGTCTGGTAGGCCGGCGCAATATGAAACGGCCTCGGTTGGTCCAAAAAACGGTCTAGCA

      TGAGATGGTGCTGCGTCAGCATCAGCGTTGTGAAGGCCGGTCCTTTACTGGCGGGCGTTG
18541 ---------+---------+---------+---------+---------+---------+ 18600
      ACTCTACCACGACGCAGTCGTAGTCGCAACACTTCCGGCCAGGAAATGACCGCCCGCAAC

      GCAGGTCAACGTGGCGCAAACGCAGTCATTGACTCCCGCCGTGGATGACCGTGTCGCGAT
18601 ---------+---------+---------+---------+---------+---------+ 18660
      CGTCCAGTTGCACCGCGTTTGCGTCAGTAACTGAGGGCGGCACCTACTGGCACAGCGCTA

      TGTCGACTATCAGCCGATTACCGGTCACCCATTGGACAAAACGCAGTTTTGGCAGACGCT
18661 ---------+---------+---------+---------+---------+---------+ 18720
      ACAGCTGATAGTCGGCTAATGGCCAGTGGGTAACCTGTTTTGCGTCAAAACCGTCTGCGA

      GGAGCAGGGGGAAGATGCGCTGCGGGAGCATAGTGCGGCGCACGTTAATGCTGAGGCGTT
18721 ---------+---------+---------+---------+---------+---------+ 18780
      CCTCGTCCCCCTTCTACGCGACGCCCTCGTATCACGCCGCGTGCAATTACGACTCCGCAA

      TGTGCGCACCACCCAACAAAAACTCAGTACCTATATTCATCGTACGATGAGTTTCCCTGC
18781 ---------+---------+---------+---------+---------+---------+ 18840
      ACACGCGTGGTGGGTTGTTTTTGAGTCATGGATATAAGTAGCATGCTACTCAAAGGGACG

      GCATTCACCCAGTGACGTGGCGCTGAAAAAACCGATGATGCCGGCCAAAAAACAGCGTCT
18841 ---------+---------+---------+---------+---------+---------+ 18900
      CGTAAGTGGGTCACTGCACCGCGACTTTTTTGGCTACTACGGCCGGTTTTTTGTCGCAGA

      GGATGTGACGCAATTGTATGCGCTGAATAGCTGCCATTCCTGGTCTGAAAAAATCCGCCA
18901 ---------+---------+---------+---------+---------+---------+ 18960
      CCTACACTGCGTTAACATACGCGACTTATCGACGGTAAGGACCAGACTTTTTTAGGCGGT

      ATTCGAGCGTGTGGCGATCATCATTGCCAGTAATTTATCGTTAAGCGCGGATCGGTTACA
18961 ---------+---------+---------+---------+---------+---------+ 19020
      TAAGCTCGCACACCGCTAGTAGTAACGGTCATTAAATAGCAATTCGCGCCTAGCCAATGT

      GGCTATGAGGGCATTATGGTCCGGCCTGCCGGGTTCAGAAGGAGCAATCCCTCTGCCTGA
19021 ---------+---------+---------+---------+---------+---------+ 19080
      CCGATACTCCCGTAATACCAGGCCGGACGGCCCAAGTCTTCCTCGTTAGGGAGACGGACT

      GCTGCCTTCGATTAACCACTGGAGTTGGTACGGTGCCTGCGGCATCGGGACAGCACAGCT
19081 ---------+---------+---------+---------+---------+---------+ 19140
      CGACGGAAGCTAATTGGTGACCTCAACCATGCCACGGACGCCGTAGCCCTGTCGTGTCGA
```

```
      TTTGGCGCAATATTTTGGCATTAGCGCTGATTGCTATGCGGTGGAAGCCGCTTGTGCCAG
19141 ---------+---------+---------+---------+---------+---------+ 19200
      AAACCGCGTTATAAAACCGTAATCGCGACTAACGATACGCCACCTTCGGCGAACACGGTC

      TTCGCTGGCCGCAGTGCATGATGCGGTCAGGGCACTGCAAGCGGGACGTTATGACGCGGT
19201 ---------+---------+---------+---------+---------+---------+ 19260
      AAGCGACCGGCGTCACGTACTACGCCAGTCCCGTGACGTTCGCCCTGCAATACTGCGCCA

      GATTGTTGGCGGCATCGAAACGGCGACATTAGAGCGGGATCTGGTGCTGTGCAGCGCGCA
19261 ---------+---------+---------+---------+---------+---------+ 19320
      CTAACAACCGCCGTAGCTTTGCCGCTGTAATCTCGCCCTAGACCACGACACGTCGCGCGT

      AATGATGCTTTCGGTCTCGCGCATTCGCCCGTTTTCACAAGGCGCTGACGGGTTTACGCC
19321 ---------+---------+---------+---------+---------+---------+ 19380
      TTACTACGAAAGCCAGAGCGCGTAAGCGGGCAAAAGTGTTCCGCGACTGCCCAAATGCGG

      GGGAGATGGTGGTGGTTTTGTTATGCTGACCCATCACCCTGTACCTCGGGCGATTGCCAC
19381 ---------+---------+---------+---------+---------+---------+ 19440
      CCCTCTACCACCACCAAAACAATACGACTGGGTAGTGGGACATGGAGCCCGCTAACGGTG

      CATTGAAGCGATATCGGGTTCTTGCGATAGCTATTCCATGACGGCGCCAGATCCACTGGG
19441 ---------+---------+---------+---------+---------+---------+ 19500
      GTAACTTCGCTATAGCCCAAGAACGCTATCGATAAGGTACTGCCGCGGTCTAGGTGACCC

      ACAGGCGCTGGCTATCAAGAAAACCCTGAGTCTGACCGCCATTGATGCTCAAACGGTGCA
19501 ---------+---------+---------+---------+---------+---------+ 19560
      TGTCCGCGACCGATAGTTCTTTTGGGACTCAGACTGGCGGTAACTACGAGTTTGCCACGT

      ATACCTTGAAGCACACGGTACAGGCACTGAACTTGGGGATCGATCGGAAGTGATGTCACT
19561 ---------+---------+---------+---------+---------+---------+ 19620
      TATGGAACTTCGTGTGCCATGTCCGTGACTTGAACCCCTAGCTAGCCTTCACTACAGTGA

      GAAGTACAGTTATCACCGTGACAAGCATTCACCGCTGTATATCGGGTCGGCGAAGTACAA
19621 ---------+---------+---------+---------+---------+---------+ 19680
      CTTCATGTCAATAGTGGCACTGTTCGTAAGTGGCGACATATAGCCCAGCCGCTTCATGTT

      CTTTGGTCACTGTTTCGCCGGTGCTGGTGCGCTCAGTTTGTGTAAAGTGTTGAGCGCATT
19681 ---------+---------+---------+---------+---------+---------+ 19740
      GAAACCAGTGACAAAGCGGCCACGACCACGCGAGTCAAACACATTTCACAACTCGCGTAA

      CGAACATGAGCGGATCCCCCCCACGCCAGTGTCTGAACTCAATGTTGATTTACCCTTGGG
19741 ---------+---------+---------+---------+---------+---------+ 19800
      GCTTGTACTCGCCTAGGGGGGGTGCGGTCACAGACTTGAGTTACAACTAAATGGGAACCC

      CGATATCCCGGCGGAGGTGCCACAACAGGCCATACCCTGGCGGCTGTCTGAGGATGGTCA
19801 ---------+---------+---------+---------+---------+---------+ 19860
      GCTATAGGGCCGCCTCCACGGTGTTGTCCGGTATGGGACCGCCGACAGACTCCTACCAGT

      GCGCAAGGCGGCGATCAACGCCTTTGGCACTGGCGGTATTAACTATCACCTTGTCATCAG
19861 ---------+---------+---------+---------+---------+---------+ 19920
      CGCGTTCCGCCGCTAGTTGCGGAAACCGTGACCGCCATAATTGATAGTGGAACAGTAGTC

      ACAATCGTCTTAAAAAGGAAGTGTAGATGAAGGTTATCGTCCATAAAATTCGCCTGAAAG
19921 ---------+---------+---------+---------+---------+---------+ 19980
      TGTTAGCAGAATTTTTCCTTCACATCTACTTCCAATAGCAGGTATTTTAAGCGGACTTTC

      ATATTACTCAGCGTTCGGCATTTCGTCATTGGGTTGAAACCACGGATTATTGCGCTTGCG
19981 ---------+---------+---------+---------+---------+---------+ 20040
      TATAATGAGTCGCAAGCCGTAAAGCAGTAACCCAACTTTGGTGCCTAATAACGCGAACGC

      AGAGCCTTGACGCCGTATTGGCCTTTGAAGTGGTTGAGGTCAGTCAGGTGGCGGAAGCGC
20041 ---------+---------+---------+---------+---------+---------+ 20100
      TCTCGGAACTGCGGCATAACCGGAAACTTCACCAACTCCAGTCAGTCCACCGCCTTCGCG
```

```
      CATTTCACTTTATTGAGATTATTCACATAAGCTCGATGGACGCCTTTGCACAAGAGATGC
20101 --------+---------+---------+---------+---------+---------+ 20160
      GTAAAGTGAAATAACTCTAATAAGTGTATTCGAGCTACCTGCGGAAACGTGTTCTCTACG

      AAACCCCGCTTTTTCAAAGACTGGTGAGCCAGTTTGATCAACTGGCTGAGGTGGTTGAGG
20161 --------+---------+---------+---------+---------+---------+ 20220
      TTTGGGGCGAAAAAGTTTCTGACCACTCGGTCAAACTAGTTGACCGACTCCACCAACTCC

      AGATAGCGGGTGAGCGGATCGCCGATGGTTATCCATCATAAAGACGGCGTGCCGCCATGG
20221 --------+---------+---------+---------+---------+---------+ 20280
      TCTATCGCCCACTCGCCTAGCGGCTACCAATAGGTAGTATTTCTGCCGCACGGCGGTACC

      GTTAAGCACGGGCATCATCATCGAGGTATTACCATTTCGGTTGCCGAATGCCACACGATA
20281 --------+---------+---------+---------+---------+---------+ 20340
      CAATTCGTGCCCGTAGTAGTAGCTCCATAATGGTAAAGCCAACGGCTTACGGTGTGCTAT

      GGGGTTGCCCTACAACAACGGATGGTGAAACAGCGGGCACGTGAACTGGCGGTAGCGCTG
20341 --------+---------+---------+---------+---------+---------+ 20400
      CCCCAACGGGATGTTGTTGCCTACCACTTTGTCGCCCGTGCACTTGACCGCCATCGCGAC

      TTGAATCGGGTTTGGCAAACCGGGGATGGCGCCTCTCAGATGGCGGGTTTTACACTGACT
20401 --------+---------+---------+---------+---------+---------+ 20460
      AACTTAGCCCAAACCGTTTGGCCCCTACCGCGGAGAGTCTACCGCCCAAAATGTGACTGA

      CATCGGCCGAGTGGTCAGCCGGTGGGTGAACATCCCCTGTTGGGCGAATATGCGGTGTCG
20461 --------+---------+---------+---------+---------+---------+ 20520
      GTAGCCGGCTCACCAGTCGGCCACCCACTTGTAGGGGACAACCCGCTTATACGCCACAGC

      ATCAGCCATTCGCATTCATGGTATGCCGGGGCGACCAGCGACACTGCGGTGGGAATCGAC
20521 --------+---------+---------+---------+---------+---------+ 20580
      TAGTCGGTAAGCGTAAGTACCATACGGCCCCGCTGGTCGCTGTGACGCCACCCTTAGCTG

      CTTCAGCAGTATCGCACTTTTGGTACAACGGCGTGGCGTTACGCCTTTGCTGCCGGTGAA
20581 --------+---------+---------+---------+---------+---------+ 20640
      GAAGTCGTCATAGCGTGAAAACCATGTTGCCGCACCGCAATGCGGAAACGACGGCCACTT

      CAGCATTGGGGCCGGGTTCAGGCGTGTGCGGCCTGGGCCATTCGGGAAGCATTTTTAAAA
20641 --------+---------+---------+---------+---------+---------+ 20700
      GTCGTAACCCCGGCCCAAGTCCGCACACGCCGGACCCGGTAAGCCCTTCGTAAAAATTTT

      AGTCATGGTTGTGGACTTCCCTATCGGCTCAGCGATATCCGGATCGATTGGCATAATGCG
20701 --------+---------+---------+---------+---------+---------+ 20760
      TCAGTACCAACACCTGAAGGGATAGCCGAGTCGCTATAGGCCTAGCTAACCGTATTACGC

      TTGGTATCGCATCCTGACCAACTGGAGTGCCGGCAGTTTTGGTTGTGGTATAGCCTGGAT
20761 --------+---------+---------+---------+---------+---------+ 20820
      AACCATAGCGTAGGACTGGTTGACCTCACGGCCGTCAAAACCAACACCATATCGGACCTA

      TGGGTATGCGCACTGTGTTTTCCGCTACAGGGATCATTGATGACTGACACGGATATTGTG
20821 --------+---------+---------+---------+---------+---------+ 20880
      ACCCATACGCGTGACACAAAAGGCGATGTCCCTAGTAACTACTGACTGTGCCTATAACAC

      GTGGTTCCCGGTGGGTCTATTGAGCGACAAGGGGAATCTCATCATGGTGAATGACACATT
20881 --------+---------+---------+---------+---------+---------+ 20940
      CACCAAGGGCCACCCAGATAACTCGCTGTTCCCCTTAGAGTAGTACCACTTACTGTGTAA

      TGAACAAGCATTGCAAAACGCCATTAATATCGCACGGTTAGCGCCATCATCACATAATTG
20941 --------+---------+---------+---------+---------+---------+ 21000
      ACTTGTTCGTAACGTTTTGCGGTAATTATAGCGTGCCAATCGCGGTAGTAGTGTATTAAC

      CCAGCCTTGGTCTGTTCACTACGATGCGGCTACCCGATGTGGTGAAGTCTCGATTGATCG
21001 --------+---------+---------+---------+---------+---------+ 21060
      GGTCGGAACCAGACAAGTGATGCTACGCCGATGGGCTACACCACTTCAGAGCTAACTAGC
```

```
      GCAGCGCGCCCTCAAGGGATTACCCTCGCTGGAGCGCGAGATGCTGATGAGTTGTGGGAT
21061 ---------+---------+---------+---------+---------+---------+ 21120
      CGTCGCGCGGGAGTTCCCTAATGGGAGCGACCTCGCGCTCTACGACTACTCAACACCCTA

      CTTCTTTGAATACTTGAGTACGTTGCTCAAGCATTCAGGCTATCCGCTTGACTGGCAGTG
21121 ---------+---------+---------+---------+---------+---------+ 21180
      GAAGAAACTTATGAACTCATGCAACGAGTTCGTAAGTCCGATAGGCGAACTGACCGTCAC

      GGTGGGGGCGCGGCAGAACGGCTCGAGCGGTATGTTGATATCGTTTGCGCCGAGTGCGCC
21181 ---------+---------+---------+---------+---------+---------+ 21240
      CCACCCCCGCGCCGTCTTGCCGAGCTCGCCATACAACTATAGCAAACGCGGCTCACGCGG

      ATGTGTCGCTGATCTGGTGGCTTATCAGCAATGGGTGCAAAGGATCAGCGATCGGCATAC
21241 ---------+---------+---------+---------+---------+---------+ 21300
      TACACAGCGACTAGACCACCGAATAGTCGTTACCCACGTTTCCTAGTCGCTAGCCGTATG

      GGTCAGAACGGCCTATCAGCCGACACAGGTGAATGAACAACAACAGGCACAACTATACGC
21301 ---------+---------+---------+---------+---------+---------+ 21360
      CCAGTCTTGCCGGATAGTCGGCTGTGTCCACTTACTTGTTGTTGTCCGTGTTGATATGCG

      GTTATTTGACCGTTCACCGGTGACCTGCGACATCAAATATGGCGAAGCAACGCGTCATGA
21361 ---------+---------+---------+---------+---------+---------+ 21420
      CAATAAACTGGCAAGTGGCCACTGGACGCTGTAGTTTATACCGCTTCGTTGCGCAGTACT

      CGTCGCTTTTCTCACGGCAAACTATGCATCGCTTGATTTTGCCGATCAACAGGCCTGGCG
21421 ---------+---------+---------+---------+---------+---------+ 21480
      GCAGCGAAAAGAGTGCCGTTTGATACGTAGCGAACTAAAACGGCTAGTTGTCCGGACCGC

      GGAAACCTATCACTATATTCGGTTCAATGAGCAGCAGGCCGCCGAAGATGGGTTTTATCT
21481 ---------+---------+---------+---------+---------+---------+ 21540
      CCTTTGGATAGTGATATAAGCCAAGTTACTCGTCGTCCGGCGGCTTCTACCCAAAATAGA

      GCATCATTTGTTTGGGCCGGTATCATGCGGATTTAGGTGGTTTTTCCGGATTGCGTTTCA
21541 ---------+---------+---------+---------+---------+---------+ 21600
      CGTAGTAAACAAACCCGGCCATAGTACGCCTAAATCCACCAAAAAGGCCTAACGCAAAGT

      CCCCAGGCTAAGCTGGCTGGCAAAACGATTGCGATTGCCTGCTTCCATGGCGAAAGGGCT
21601 ---------+---------+---------+---------+---------+---------+ 21660
      GGGGTCCGATTCGACCGACCGTTTTGCTAACGCTAACGGACGAAGGTACCGCTTTCCCGA

      AGCCGAACTGGTGGTGGAAGGCCCGCAATACCTGGCGTTGAGTCTTGAACATGAGAGCGA
21661 ---------+---------+---------+---------+---------+---------+ 21720
      TCGGCTTGACCACCACCTTCCGGGCGTTATGGACCGCAACTCAGAACTTGTACTCTCGCT

      CGAAAATCTGTTTATCGCCGGAATGAAGCTTGGCCAACTGTGGCTGATGTTACAGTCGTG
21721 ---------+---------+---------+---------+---------+---------+ 21780
      GCTTTTAGACAAATAGCGGCCTTACTTCGAACCGGTTGACACCGACTACAATGTCAGCAC

      GGGGTGGAGCTTGCATCCGCTCAGTGTGCTGGTGCAGCACGCCACTGCGCGGTGCGCTTT
21781 ---------+---------+---------+---------+---------+---------+ 21840
      CCCCACCTCGAACGTAGGCGAGTCACACGACCACGTCGTGCGGTGACGCGCCACGCGAAA

      GGCTGACACTGTCCGGCTTACCGGCTTACCGGTTTTTTTTGCGCGTTTTGGCCAGCACCG
21841 ---------+---------+---------+---------+---------+---------+ 21900
      CCGACTGTGACAGGCCGAATGGCCGAATGGCCAAAAAAAACGCGCAAAACCGGTCGTGGC

      TCAGTCGGGAATACCCACTCCAAGACGCGCCTGGCAGCGTATTTTAACCACAACACAGCA
21901 ---------+---------+---------+---------+---------+---------+ 21960
      AGTCAGCCCTTATGGGTGAGGTTCTGCGCGGACCGTCGCATAAAATTGGTGTTGTGTCGT

      TAGTTTCTCCCCTGAAAACGGGGCTGATGTAAAACAACCCTAACCGGATAAGGAATCGAT
21961 ---------+---------+---------+---------+---------+---------+ 22020
      ATCAAAGAGGGGACTTTTGCCCCGACTACATTTTGTTGGGATTGGCCTATTCCTTAGCTA
```

```
      AATGGACGTTTGGATTGCAGTGGTTGTCGTATTATTAGTCATGTCTTTTCTGGCCTTCGT
22021 ---------+---------+---------+---------+---------+---------+ 22080
      TTACCTGCAAACCTAACGTCACCAACAGCATAATAATCAGTACAGAAAAGACCGGAAGCA

      GACGAGCCAAAAGGCACGGGATACCCGGCCGATGTTTTTGTGGGGATTCAATGCGATGGG
22081 ---------+---------+---------+---------+---------+---------+ 22140
      CTGCTCGGTTTTCCGTGCCCTATGGGCCGGCTACAAAAACACCCCTAAGTTACGCTACCC

      GCCGGTTGCCTTGGTATTTTGCTATTTTGGTGGGGGTGATATAAACCATAAAGCTCTGAT
22141 ---------+---------+---------+---------+---------+---------+ 22200
      CGGCCAACGGAACCATAAAACGATAAAACCACCCCCACTATATTTGGTATTTCGAGACTA

      CCTGGCCTTTGTCGGTATTTATCTACTTCATATGAATATTGTACTGACACTCTGGTACGG
22201 ---------+---------+---------+---------+---------+---------+ 22260
      GGACCGGAAACAGCCATAAATAGATGAAGTATACTTATAACATGACTGTGAGACCATGCC

      CAATACTGCTGCATCCAAACTCAAGGATGTGTTACCCACACAACAGGTGCCGTTATTGGC
22261 ---------+---------+---------+---------+---------+---------+ 22320
      GTTATGACGACGTAGGTTTGAGTTCCTACACAATGGGTGTGTTGTCCACGGCAATAACCG

      GGTGATGATGGTGAACATCTTCGGGTGGTTGTATTGCCTGCCTTTTTACTGGGCCAGTCA
22321 ---------+---------+---------+---------+---------+---------+ 22380
      CCACTACTACCACTTGTAGAAGCCCACCAACATAACGGACGGAAAAATGACCCGGTCAGT

      ATTGCAAGGGCCATTTGGCACAGTTCAATGGGTAGCCATCGTGGTTTATAGTCTCGGCAC
22381 ---------+---------+---------+---------+---------+---------+ 22440
      TAACGTTCCCGGTAAACCGTGTCAAGTTACCCATCGGTAGCACCAAATATCAGAGCCGTG

      CCTCTATCACTTTGGCGGCGACTACCAAAAGCGCCGCTTCAAACAAAATCCGCAAAATAA
22441 ---------+---------+---------+---------+---------+---------+ 22500
      GGAGATAGTGAAACCGCCGCTGATGGTTTTCGCGGCGAAGTTTGTTTTAGGCGTTTTATT

      AGGCCAGATCCTCTGCTCCGGGTTTTGGGCGACGTCTCGTCATCCCAATTATTTTGGCGA
22501 ---------+---------+---------+---------+---------+---------+ 22560
      TCCGGTCTAGGAGACGAGGCCCAAAACCCGCTGCAGAGCAGTAGGGTTAATAAAACCGCT

      CTTTCTGATTTTTGCCAGTTTCGGACTGCTGGCGGGCAACTGGTTTGGGGTAATTGCGCC
22561 ---------+---------+---------+---------+---------+---------+ 22620
      GAAAGACTAAAAACGGTCAAAGCCTGACGACCGCCCGTTGACCAAACCCCATTAACGCGG

      GCTGACCAATATTGTCCAATACTTCGCTGATGCGATTCCCAAAAGTGAGAAAATGGCCGA
22621 ---------+---------+---------+---------+---------+---------+ 22680
      CGACTGGTTATAACAGGTTATGAAGCGACTACGCTAAGGGTTTTCACTCTTTTACCGGCT

      ACAGCGCTATGGTAACGCCTGGCTTGACTATAAAAGGAAAGTGAAGTGCTTTATCCCCTT
22681 ---------+---------+---------+---------+---------+---------+ 22740
      TGTCGCGATACCATTGCGGACCGAACTGATATTTTCCTTTCACTTCACGAAATAGGGGAA

      TATCTTTTAGTGGATAGCTCAGGGAACAGTAGCGGTTTGATTAACGATAGCCTGGGGCAT
22741 ---------+---------+---------+---------+---------+---------+ 22800
      ATAGAAAATCACCTATCGAGTCCCTTGTCATCGCCAAACTAATTGCTATCGGACCCCGTA

      CTAAGGTCACTCTTTCATGGCATATGCTCTATTCAAGCGGTAAAATATTGCCACCATCTA
22801 ---------+---------+---------+---------+---------+---------+ 22860
      GATTCCAGTGAGAAAGTACCGTATACGAGATAAGTTCGCCATTTTATAACGGTGGTAGAT

      CTGTTAATGAGCATCAAATGATTAAGTTTTCTGAAATCGGCAAGGCTATGACAGAGTTCG
22861 ---------+---------+---------+---------+---------+---------+ 22920
      GACAATTACTCGTAGTTTACTAATTCAAAAGACTTTAGCCGTTCCGATACTGTCTCAAGC

      AATCCGTCGATGGCGTTCCATTACAAGGATATGGTATCTCCGACTCTCGTGGGACGATCG
22921 ---------+---------+---------+---------+---------+---------+ 22980
      TTAGGCAGCTACCGCAAGGTAATGTTCCTATACCATAGAGGCTGAGAGCACCCTGCTAGC
```

```
      TGCTTTTTGATGATTACGCCGAAGGATTACAAGATATTGAGGGGTTCTCCCACCTTTATT
22981 --------+---------+---------+---------+---------+---------+ 23040
      ACGAAAAACTACTAATGCGGCTTCCTAATGTTCTATAACTCCCCAAGAGGGTGGAAATAA

      TGCTGTTTCATCTGCATCGTTCTAATAACCATAGTCTTAAGGTAAAACCGTTCCTGGATA
23041 --------+---------+---------+---------+---------+---------+ 23100
      ACGACAAAGTAGACGTAGCAAGATTATTGGTATCAGAATTCCATTTTGGCAAGGACCTAT

      CTCAGGAGCGAGGCGTTTTTGCTACTCGCTCACCAAGACGGCCTAATCCACTTGGTCTCA
23101 --------+---------+---------+---------+---------+---------+ 23160
      GAGTCCTCGCTCCGCAAAAACGATGAGCGAGTGGTTCTGCCGGATTAGGTGAACCAGAGT

      CTATTGTGCAATTAGAACGCCGGGAGGGAAATGTCTTGCATGTTCGAGGTCTGGATCTGC
23161 --------+---------+---------+---------+---------+---------+ 23220
      GATAACACGTTAATCTTGCGGCCCTCCCTTTACAGAACGTACAAGCTCCAGACCTAGACG

      TGAACGGCACACCCATTATTGATATCAAACCTTATATTCAAGCTTTCGATAATGTTCAGG
23221 --------+---------+---------+---------+---------+---------+ 23280
      ACTTGCCGTGTGGGTAATAACTATAGTTTGGAATATAAGTTCGAAAGCTATTACAAGTCC

      GAACCCGTGATGGCTGGTATTTGCAGGGGATGAATCCACATCAGACGCTATCTGACGAGC
23281 --------+---------+---------+---------+---------+---------+ 23340
      CTTGGGCACTACCGACCATAAACGTCCCCTACTTAGGTGTAGTCTGCGATAGACTGCTCG

      GTTTTAAATAGAACATGAATTCGTTACCGGTTGCTGTTTTCATTTTTATTGCCGAGGTTA
23341 --------+---------+---------+---------+---------+---------+ 23400
      CAAAATTTATCTTGTACTTAAGCAATGGCCAACGACAAAAGTAAAAATAACGGCTCCAAT

      AGATGCCGCCGGTTGGTTTTGTGGCTATGTTGTGTATGGTGTCAGCGCCCACAAAAGCAG
23401 --------+---------+---------+---------+---------+---------+ 23460
      TCTACGGCGGCCAACCAAAACACCGATACAACACATACCACAGTCGCGGGTGTTTTCGTC

      TAGTGATCCAGATTCAAGATCTTTGCTGCTGCTGAGATCGTGACACCAAAAAAGGTAACC
23461 --------+---------+---------+---------+---------+---------+ 23520
      ATCACTAGGTCTAAGTTCTAGAAACGACGACGACTCTAGCACTGTGGTTTTTTCCATTGG

      GGTTTTTGAGATTGTAAAAATTGGCAGATGGTGCCATTAACCAACGTCGATCCGGTTGCC
23521 --------+---------+---------+---------+---------+---------+ 23580
      CCAAAAACTCTAACATTTTTAACCGTCTACCACGGTAATTGGTTGCAGCTAGGCCAACGG

      AGGATCTGGTCACACCAATCGATGGCTGTCTCGGTCTGCGATTCATCTTCGATGGTGACG
23581 --------+---------+---------+---------+---------+---------+ 23640
      TCCTAGACCAGTGTGGTTAGCTACCGACAGAGCCAGACGCTAAGTAGAAGCTACCACTGC

      CCATATTTGTGCTGGCCTATATTGCTCGGATCAAGATCGATTACCCGAATAGGCTGTTTG
23641 --------+---------+---------+---------+---------+---------+ 23700
      GGTATAAACACGACCGGATATAACGAGCCTAGTTCTAGCTAATGGGCTTATCCGACAAAC

      CTCGACAAGTTTTCCAGAATCCTCGGCTGTAGACCGATCAGTAAAATTTTTTGATCTGCC
23701 --------+---------+---------+---------+---------+---------+ 23760
      GAGCTGTTCAAAAGGTCTTAGGAGCCGACATCTGGCTAGTCATTTTAAAAAACTAGACGG

      ATCACCGCACTATCTACCTGTTTGCCGCACAGGACGATTTCATCGTTTTTACAATGTACG
23761 --------+---------+---------+---------+---------+---------+ 23820
      TAGTGGCGTGATAGATGGACAAACGGCGTGTCCTGCTAAAGTAGCAAAAATGTTACATGC

      GTCTTATCACATAGTTTCAGGTGTCGATAAACTGCATTGAGTGCGGCAATAAAATCAGCC
23821 --------+---------+---------+---------+---------+---------+ 23880
      CAGAATAGTGTATCAAAGTCCACAGCTATTTGACGTAACTCACGCCGTTATTTTAGTCGG

      CGCTGCTGGTCATTCTCCAGTGGCAGTTCAAGGATCTGCTGTAGTGAATACTCGCTATTG
23881 --------+---------+---------+---------+---------+---------+ 23940
      GCGACGACCAGTAAGAGGTCACCGTCAAGTTCCTAGACGACATCACTTATGAGCGATAAC
```

```
      TGATAGTGATCGGTAAACGCCTGGCCCCGGCTCCCCATAAAATCCGCCTCGATCATCTTT
23941 ---------+---------+---------+---------+---------+---------+ 24000
      ACTATCACTAGCCATTTGCGGACCGGGGCCGAGGGGTATTTTAGGCGGAGCTAGTAGAAA

      TCATGACCTCGAATAATGGGGTAGTCCTGATGTTGAGGATGACCAATAGCTTCAACCGAG
24001 ---------+---------+---------+---------+---------+---------+ 24060
      AGTACTGGAGCTTATTACCCCATCAGGACTACAACTCCTACTGGTTATCGAAGTTGGCTC

      CTGAGCGCCTGACAGCGGATTTGCAACGTTTCCCGCTGAAGATCGAATTTGTTAACTTGC
24061 ---------+---------+---------+---------+---------+---------+ 24120
      GACTCGCGGACTGTCGCCTAAACGTTGCAAAGGGCGACTTCTAGCTTAAACAATTGAACG

      TCAAGAAAATCCTTTTTGAGAAGCGCATACAGCTTGTTGATTTCGTTATCTGAATGCAGC
24121 ---------+---------+---------+---------+---------+---------+ 24180
      AGTTCTTTTAGGAAAAACTCTTCGCGTATGTCGAACAACTAAAGCAATAGACTTACGTCG

      TTGTTGTGCTGATAGGATGATTCGAGTTGTTGTAGCTGTTGTATAAATTGCGGGTAATTT
24181 ---------+---------+---------+---------+---------+---------+ 24240
      AACAACACGACTATCCTACTAAGCTCAACAACATCGACAACATATTTAACGCCCATTAAA

      AGTTTTGTTGCGTCAGAAGACCACATCATCTTCCAAATGCTACGCTTTTTTAGGCCCAAT
24241 ---------+---------+---------+---------+---------+---------+ 24300
      TCAAAACAACGCAGTCTTCTGGTGTAGTAGAAGGTTTACGATGCGAAAAAATCCGGGTTA

      TTACGCAATGTTATGATCTTTCTGAGGTCATGGTAATTCATAGATATCTCTTTATACTCT
24301 ---------+---------+---------+---------+---------+---------+ 24360
      AATGCGTTACAATACTAGAAAGACTCCAGTACCATTAAGTATCTATAGAGAAATATGAGA

      GGTCCGGATGAGCTGCGGGCTTGTCCAGACATGTGTTTTCCCGTGAGTCAGACAGTCAAT
24361 ---------+---------+---------+---------+---------+---------+ 24420
      CCAGGCCTACTCGACGCCCGAACAGGTCTGTACACAAAAGGGCACTCAGTCTGTCAGTTA

      CACTGTACATCAGACGGTTTCTAGCGTGTTGGCTAATGTCCGGCGTGAAGTCCCTGTTGA
24421 ---------+---------+---------+---------+---------+---------+ 24480
      GTGACATGTAGTCTGCCAAAGATCGCACAACCGATTACAGGCCGCACTTCAGGGACAACT

      TATCGTAGCCCGTCGTTGTTACATCCTCGGGTAAATCAGCTATGTCGACTATTCGAGCTC
24481 ---------+---------+---------+---------+---------+---------+ 24540
      ATAGCATCGGGCAGCAACAATGTAGGAGCCCATTTAGTCGATACAGCTGATAAGCTCGAG

      ATCTGTATAAATTCCTTTTTGACACAGAGATCCGAACTTTGATTCAGTTTAACGCTTCGG
24541 ---------+---------+---------+---------+---------+---------+ 24600
      TAGACATATTTAAGGAAAAACTGTGTCTCTAGGCTTGAAACTAAGTCAAATTGCGAAGCC

      CAGGCTGCCTGGTCGCAAAGTTTCTCTTTATTTTAAGGTTTGGAAACGATCCTGCTGATC
24601 ---------+---------+---------+---------+---------+---------+ 24660
      GTCCGACGGACCAGCGTTTCAAAGAGAAATAAAATTCCAAACCTTTGCTAGGACGACTAG

      AATCAAAAGAGCAGGTGCTACCTATAGAACAATATGGCATATGCTATATTTAAGCAAGAA
24661 ---------+---------+---------+---------+---------+---------+ 24720
      TTAGTTTTCTCGTCCACGATGGATATCTTGTTATACCGTATACGATATAAATTCGTTCTT

      GAACCTCCCCCTAATCTGGTCGATCTGAAGTACGCTTTACTCTGACGTGGTTAGGCCGGC
24721 ---------+---------+---------+---------+---------+---------+ 24780
      CTTGGAGGGGGATTAGACCAGCTAGACTTCATGCGAAATGAGACTGCACCAATCCGGCCG

      TACCCGCCGTCTTGTCTGGAAACGATTTTGCTCTGTTATCGTACTTTTATTTCCCTTCCC
24781 ---------+---------+---------+---------+---------+---------+ 24840
      ATGGGCGGCAGAACAGACCTTTGCTAAAACGAGACAATAGCATGAAAATAAAGGGAAGGG

      CTCCTGATAATTGTTCATAGATAGTTAACGTTGTGTTGCGTGATAGCCTGGCTGTCGCAT
24841 ---------+---------+---------+---------+---------+---------+ 24900
      GAGGACTATTAACAAGTATCTATCAATTGCAACACAACGCACTATCGGACCGACAGCGTA
```

```
       TATGCCCCGCGAGACGATTGACAGCTGAAAGGTGACAGCTAGTACATCGTAATACTGATG
24901  --------+---------+---------+---------+---------+---------+ 24960
       ATACGGGGCGCTCTGCTAACTGTCGACTTTCCACTGTCGATCATGTAGCATTATGACTAC

       ATGTATTTGTTATATCACTTATTATTCATGCGTTTATTAGCTATTCTATCAAGATGCGCA
24961  --------+---------+---------+---------+---------+---------+ 25020
       TACATAAACAATATAGTGAATAATAAGTACGCAAATAATCGATAAGATAGTTCTACGCGT

       ACAACAGATCATCATCTCACGATCGAAATTGGATCC
25021  --------+---------+---------+------ 25056
       TGTTGTCTAGTAGTAGAGTGCTAGCTTTAACCTAGG
```

IV(ii) PUTATIVE PROTEINS ENCODED BY THE 5' END OF THE *pig* CLUSTER

Listed below are the predicted amino acid sequences of putative proteins, together with predicted MW and pI. Amino acids are denoted by single letter symbols.

Orf(-1)

...AEWVTGNEMLVPLAGPANRQGRIAAENMLGGERRYQRSQGTAICKVFDHAVGSVGLNEKALGRLQMDYQ
KVYIHTPNHASYYPGSSPISFKLLFNPITGDIFGAQAVGKAGIDKCIDVIAVAQRAHLKVQDLEYLELAYAP
PFNSARDVVNQAGMLASNVINGDTKICHVVDIVNIDHQGQCLLDIRTSQELKTIGTYPEALHIPVDELRGRL
NELPKDKEILIGCQSGLRGHVAYRLLTQHGFRARNLSGGYKTYSLTLSA

MW= 2861.8 Da; pI= 8.3

Orf1

MEAMECFMDFNLSNSQSDIYESAYRFACDVLDQDAQTRISQKILSTELWKKAAAYGFAHGPVSHQFGGSELG
ALDTALMIEALGKGSRDIGLSFSLCAHLCACVIPLYRFGSSELKDKYLESLVTGKLIAANAATEPDAGSDIY
NMQATAQPCEGGYILNGKKIFITNAPIADVFIIYAKTNPDHGFLGVSAFLIEKGTPGLNVGEVIPKDCLSNC
PWSEIVFNDIFIPQSQRIGMEGAGGAIFHDSMIWEKGCLSALFVGGLARLLETTLEYAKARQQFGKAIGQFQ
SVSNRIIDMKLRLEQCRLMLYRACWKHDQGQDAEADIAMSKLLISEYAVQSGLDAIQTFGGAAMDQELGLVR
HLLNMIPSRIFSGTNDIQKEIIARKLGLRGTSS

MW= 4289.2 Da; pI= 5.4

Orf2

MIIQRLFGILYMLAGLAKAFPQFENVPAVLRQAAIANQGTWYAAASIWLGAHGDVINILVGVVLFGSGVILM
LNPLWTTLVIYAQLLMMAVFVVILHQSQPQVMLLDGVFALAALYMLRGQYHRKPKPRTFPTTSFSLPTPSSE
SSFSAPLGDEYDVVIIGGGASGLTAASEFTHERVLVLEKSSTFGGNARYHTFNRLKHPTAGVCFQEPFPGSN
MLRLLKKIGLEGKYKSNEKDTLVFFDTFLLLKCLGEIVVGFIKQPRYLLKLSVWGLTSQLFLHAIIGKPYVV
AAKQLGDPIFADLYTFLDKFSPRGDFYPRLPWTPNGSWSKAHMELLDNISLYTYLFEPDKLGRLPEQLRPPA
RLGKLVENAVSTTLRVECLDIHDVSAYVGLHFLVGYLRGNLVTLPGGNGSISAGLCKYLSHQRNVTLQNHVQ
LTAVEPQHNGTCIQFTINGQPRQVQAQQIIWAAPKTQLATWLPGLPAKQLAAIKNIRHEDYYLANVFLSKPV
LGHSFGGYMIEPDSNKDPFSWCKAGTCLVANWMDDHADVDVGVLTLLKPTTRSERQDRTAQNAFLALQQQTY
AEIAKVLRNIGIGAEVIEDIQIWYWPAGLVTSVVGQQAEGVFETARQSFENIHFANQDSVGVGNIESAILSG
IDAANAVKAQLMDTENVVEVAG

MW= 73860.3 Da; pI= 7.27

Orf3

MNQPLVVEISGDKALEHHHLGGKGYSLNNLIHAGLPVPSAFCVTAQAYQQFIEEVVPGAELTDGDLIAVRDA
ILHADIPDSLKQAIGDAYQHLGHDTTIAVRSSALDEDGQRQSFAGQYETYLHVKGSEAVLHKVQACWASLWA
ERAAQYRHESASHSAIAVILQVMVDADAAGVMFTQDPLSGSTDKVVIDSCWGLGEGVVSGQVTTDSFTLDKA
TGELCDQQIRHKPNYCQRDEHGLVTLLQTPEAKRDLPSLTPAQLQQLVTLARQAQLIYSTELDIEWAVKDDK
VWLLQARPVTTSAKTANVIYANPWESDPAAKEGAFFSRMDTGEIVTGLMTPLGLSFCQFYQKHIHGPAIKTM
GLADISHWQIYMGYIQGYVYLNISGSAYMLRQCPPTRNEMKFTTRYATDEIDFKDYKNPYGAGVQGWDYAKS
CWYWLKQQVRNMRSAARTVEQMIALRQDETTRFLGLDLTAMTLQQLDQELQRIDRFFLDSCAAYMPFFLQSF
ALYDALAQACERHIKDGKGLQNRIKASMNNLRTIEVTLGIIKLVATVNQQTELKALFEQHRADELVTLLPVH
DISRAFWQGDFEDFLVEFGSRGRQEFDLSIPRWRDDPSYLLQVMKMYLQHPVDLHKKLRETELLRQQDSEAL
FSAMSWSGRFKLKTLIKLYGMMAERREATRPTFITETWFYRCIMLEVLRRLDAQGIASSADLPYVDFEQFRA
YVAGTIPAEQAFSKARLDQNRHQHLFNLHAEEPPMAIVGPYTPKVKAPTQDDKTIRSLTGLAASPGNVVAKA
RVITDLQVQAGEFQPDEILVARFTDASWTPLFALAAGIVTDIGSTLSHSCIVAREFGIPAVVNLKTATQIIN
SGDMLILDGDSGTVIIQHQEERNHDG

MW= 99766.4 Da; pI= 5.59

Orf4

MTTMIGQTRQAGSSSYEQAWQAEQAPCPGMEPDTLTVGVVVVTRNPTFFQTGLSVLNDIRDYVFNRVHIQSE
LPLKLSELASDPLYSEAREKAIHFLKNQSKALNIQVIQCASLAEATGKIIYTHALEQQPEFQMGMLFYDQTS
LGNVDDSIEKIDRDLDAFYSAMQRGGIPAFYTTFSTVTFIRDVRSSFRYLPQQYREIVRSEDPAIFQTELLC
LWMDFFEMNYTNRRVKPIGALALHNTLAEQLIQFFERTAASRWLVSYYTGSIISNLIGYLDRHAEAHGALVL
RGPNEHAIACGAMANWQLYRMPFLGVVTSGMMDEFKGTLINLKETAAQGIIVAAENRNNQWYSFQGTQTPTE
DMRDVLAAKRIPYVYIDDVDGIADGLAEVFRLYHQAQGPVVILATQNVLESTLSLEPVPGDLPPVSGLPAYD
CPPISDSFEQAMALINEGPEKLVWQLGPVSDDEYALVHEIADAAGLALVDSLAHPGSAPKYYQGKRNPHYLG
TLAIYGYSPRVYNFLHTNDKLNPMSDQSVFMIKSRVAQITTPFSDGRLERKVHLVQLTHDERHLSPYADLKL
HMDCLTFLRAVKANLHVDAALREKRKALIAAYLDSPSDVVSQLPSLPMSANYFFCQLNRVIENLIKTENFDF
TGVYDVGRCGISAVRNVAKTRRGFSGWYGRALMGDALLATSYLAHTSPTHVVAFIGDGAKGIVPDILPAFID
NILTHPQLLNKSITIFYFCNGGLSVINTYQERILFNRTSRQMRLVNVDQPAFEQTVDDFHIQGKTLTHFDED
TIRHALMTPKRLNLFSVVLGHNNEGDGISLATAKGWQRDPSDREALQERKDWAARQPESTSTSFDQGQNKEA
IS

MW=96977.1 Da; pI=5.66

Orf5

MKFGFIAHPTSVGLKRYVKMIDLLQRNSTELHSGYKRDLWRRENLVPFMNFAKITSATGATCEGVIKYMPLV
ADEMLADARGIANRVVSGIEELVEDGAELVGLGGFTSIVGRRGEATAEKSPVPVTSGNSLTTYAGYKALMQI
QSWLDIQPEQEPVAIVGYPGSICLALSRLLLAQGFSLHLLHRAGHKDEDELLSHLPEQYRSRVTLTSDPEDL
YPRCKLFVAATSAGGVIDPYKLQPGSVFIDVALPRDINSDTRPDRDDILIIDGGCVTATDAVKLGGESLNVT
IKQQLNGCMAETIVLALENRRENFSLGRYLALDNVLEIGELAEKHGFLVYPLASYGERIDRQRVINLKRYYH
HDIYSDEPDTEQPPASQLAFIDAIIAQDPAREDTLDRYHQFINPMMVEFLKLQHCDNVFRRASGTQLFTADG
EAFLDMVAGYGCINLGHNPQPIIDALKAYLDAQGPNFIQYISIPEQAAKLAEVLCHFAPGNMGRVFFSNSGT
EAVEAAMKLAKASTGKAGIAYLKNSYHGKTLGALSITGREKHRRHFKPLLASMIEVPFADIEALRQTLSRDD
IGALMIEPIQGEGGVHVPPPGYLRTVQEICRQTDTLLMVDEVQTGLGRTGKLFACEWEGIEPDVLMLSKSLS
GGVMPIGATLCRAIFGNGPYGTADRFLMHSSTFGGGNIAAVVALSALREILAQDLVGNAERLGTYFKQALTD
VAARYPFVAEIAGRGLMLGIQFDQTFAGAVGASAREFATRLPGDWHTTWKFLPDPVQAHLKAAMERMEQSLG
EMFCMKFVTKLCQDHNILTFITANSSTVIRIQPPLTISKAEIDRFVSAFATVCDELSTFLE

MW= 93577.3 Da; pI= 5.6

REFERENCES

Ahmed, S.A. and Clairborne, A. (1989) The streptococcal flavoprotein NADH oxidase I. Evidence linking NADH oxidase and NADH peroxidase cysteinyl redox centers. *J Biol Chem* **264(33):** 19856-19863

Alberti, L., and Harshey, R.M. (1990) Differentiation of *Serratia marcescens* 274 into Swimmer and Swarmer cells. *J Bacteriol.* **172(8):** 4322-4328

Aldea M, Garrido T, Hernandez-Chico C, Vicente M, Kushner SR (1989) Induction of a growth-phase-dependent promoter triggers transcription of *bolA*, an *Escherichia coli* morphogene. *EMBO J* **8(12):**3923-3931

Almirón M, Link AJ, Furlong D, Kolter R (1992) A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev* **6(12B):**2646-2654

Altschul, S.F., Gish, W., Miller, W., Myers, E.W and llipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403-419

Arai, M., Haneishi, T., Kitihara, N., Enokita, R. and Kawakubo, K. (1976) Herbicidins A and B, two new antibiotics with herbicidal activity. Producing organism and biological activities. *J Antibiot* (Tokyo) **29(9):** 863-869

Armstrong, G.A. (1994) Eubacteria show their true colours: genetics of pigment biosynthesis from plants to microbes. *J Bacteriol* **176(16):** 4795-4802

Axelrod, P.E., Rella, M. and Schroth, M.N. (1988) Role of antibiosis in competition of *Erwinia* strains in potato infection courts. *Appl Env Microbiol* **54:** 1222-1224

Bainton, N.J., Stead, P., Chhabra, S.R., Bycroft, B.W., Salmond, G.P.C., Stewart, G.S.A.B. and Williams, P. (1992) *N*-(3-oxohexanoyl)-L-homoserine lactone regulates carbapenem antibiotic production in *Erwinia carotovora*. *J Biochem* **288:** 997-1004

Bankier A.T., Weston, K.M. and Barrel B.G. (1987) Random cloning and sequencing by the M13 dideoxynucleotide chain termination method. *Meth Enzymol* **155:** 51-93

Battaile K.P., Mohsen, A.W. and Vockley, J. (1996) Function of the active-site glutamate-368 in rat short-chain acyl-CoA dehydrogenase. *Biochemistry* **35(48):** 15356-15363

Becker D.F., Fuchs, J.A., Banfield, D.K., Funk, W.D., MacGillivray, T.A., and Stankovitch, M.T. (1993) Characterization of the wild-type and an active-site mutant in *Escherichia coli* of Short-chain acyl-CoA dehydrogenase from *Megasphaera elsdenii*. *Biochem* **32:** 10736-10742

Berg D.E and Berg, C.M. (1983) The prokaryotic transposable element *Tn5*. *Biotechnol* **1:** 417-435

Berg, D.E. (1977) The prokaryotic transposable element Tn*5*. *Biotechnol* **1:** 417-435

Berg, D.E. (1983) Structural requirement for IS*50* mediated gene transposition. *Proc Natl Acad Sci* USA **80:** 792-796

Berg, D.E., Weiss, A. and Crossland, L. (1980) Polarity of Tn*5* insertion mutations in *E. coli*. *J Bacteriol* **142:** 439-446

Bibb, M., Biró, S., Motamedi, H., Collins, J.F. and Hutchinson, C.R. (1989) Analysis of the nucleotide sequence of the *Streptomyces glaucescens tmcI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *EMBO J* 8(9): 2727-2736

Bochner, B.R., Huang, H.C., Shieven, G.L. and Ames, B.N. (1980) Selection for the loss of tetracycline resistance. *J Bacteriol* 143: 926-933

Brandsch, R., Bichler, V. and Nagursky, H. (1987) Covalent flavinylation of 6-hydroxy-D-nicotine oxidase analyzed by partial deletions of the gene. *Eur J Biochem* 165(3): 559-564

Braun, V. and Schmidtz (1980) Excretion of protease by *Serratia marcescens*. *Arch Microbiol* 124: 55-61

Broome-Smith, J.K., Taddyon, M. and Zhang, Y. (1990) Beta-lactamase as a probe of membrane protein assembly and protein export. *Mol Microbiol* 10: 1637-1644

Bu'Lock, J.D. (1961) Intermediary metabolism and antibiotic synthesis. *Annu Rev Microbiol.* 3:293-342

Buchan, T., Roach, C., Ruby, C.and Taylor, D. (1994) Mutants of *Streptomyces cattaleya* defective in synthesis of A-factor required for thienamycin production. J *Antibiot* 47(9):992-1000

Bult C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton,. G.G., Blake J.A., FitzGerald L.M., Clayton, R.A., Gocayne, J.D., Kervalage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick J.M., Glodek, A., Scott, J.L. Geoghagen, N.S.M. (1996) Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii. Science* 273(5278): 1058-1073

Burger, S.R., and Bennet, J.W. (1985) Droplet Enrichment Factors of Pigmented and Nonpigmented *Serratia marcescens* : Possible Selective Function for Prodigiosin. *Appl Envir Microbiol.* 50(2): 487-490

Bycroft, B.W., Maslen, C., Box, S.J., Brown, A. and Tyler, J.W. (1987) The isolation and characterisation of (3R,5R)- and (3S,5R)-carbapenem-3-carboxylic acid from *Serratia* and *Erwinia* species and their putative biosynthetic role. *J Chem Soc Chem Commun* 21: 1623-1625.

Calogero, S., Gardan, R., Glaser, P., Schweitzer, J., Rapoport, G. and Débarbrouillé, M. (1994) RocR, a novel regulatory protein controlling arginine utilization in *Bacillus subtilis*, belongs to the Ntrc/NifA family of transcriptional activators activators. *J Bacteriol* 176: 1234-1241

Carter , P., Bedouelle, H. and Winter, G. (1985) Improved oligonucleotide site-directed mutagenesis using M13 vectors. *Nuc Acids Res* 13: 4431-4443

Chang, A.C.Y. and Cohen, S.N. (1978) Construction and characterisation of an amplifiable multicopy DNA cloning vehicle derived from the P15A cryptic miniplasmid. *J Bacteriol* 134: 1141-1156

Chhabra, S.R., Stead, P., Bainton, N.J., Salmond, G.P.C., Stewart, G.S.A.B., Williams, P. and Bycroft, B. (1992) Autoregulation of carbapenem

biosynthesis in *Erwinia carotovora* by analogues of N(-3-oxohexanoyl-)L-homoserine lactone. *J Antibiot* **16(3): 441-443**

Chou, P. and Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequences. *Adv Enzymol* **47: 45-148**

Cicicopol C., Peters, J., Kellermann, J. and Baumeister, W. (1994) Primary structure of a multimeric protein, homologous to the PEP-utilizing enzyme family and isolated from a hyperthermophilic archaebacterium. *FEBS Lett* **356: 345-350**

Clegg, S. and Allen, B.L. (1985) Molecular cloning and expression of of an extracellular nuclease of *Serratia marcescens* in *Escherichia coli. FEMS Microbiol Lett* **27: 257-262**

Clement, J.M., Perrin, D. and Hedgbeth, J. (1982) Analysis of receptor and beta-lactamase synthesis and export using genes in a minicell system. *Mol Gen Genet* **185: 304-310**

Clements-Jewry, S. (1976) The reversal of glucose-repressed prodigiosin production in *Serratia marcescens* by the cyclic 3'5'-adenosine monophosphate inhibitor theophylline. *Experientia* **32(4): 421-422**

Coco, E.A., Narva, K.E., and Feitelson, J.S. (1991) New classes of *Streptomyces coelicolor* A3(2) mutants blocked in undecylprodigiosin (Red) biosynthesis. *Mol Gen Genet* **227: 28-32.**

Coderre P.E. and Earhart, C.F.(1989) The *entD* gene of the *Escherichia coli* K12 enterobactin gene cluster*J Gen Microbiol* **135(11):3043-3055**

Comeau, D.E. Ikenaka, K., Tsung, K. and Inyoue, M. (1985) Primary characterization of the protein productd of the *Escherichia coli ompB* locus: structure and regulation of the synthesis of the OmpR and EnvZ proteins. *J Bacteriol* **164: 578-584**

Cox .A.R.J.(1995). Molecular genetic investigation into the biosynthesis and regulation of antibiotic (5R)-Carbapen-2em-3-carboxylic acid, by *Serratia marcescens* and *Erwinia carotovora* subspecies *carotovora.* PhD Thesis, Warwick University.

Cox, A.R., Thomson, N.R., Bycroft, B., Stewart G.S., Williams, P., Salmond, G.P. (1998) A pheromone-independent CarR protein controls carbapenem antibiotic synthesis in the opportunistic human pathogen *Serratia Marcescens. Microbiology* **144 (1): 201-209**

Craig, E.A, Gambill, B.D. and Nelson, R.J. (1993) Heat shock proteins: Molecular chaperones of protein biogenesis. *Microbiol Rev* **57(2): 402-414**

Craig, E.A. and Gross, C.A. (1991) Is Hsp70 the cellular thermometer? *Trends Biochem Sci* **16(4): 135-140**

Cullen, J.C. (1994) The Miracle of Bolsena: The growth of *Serratia* on sacramental bread and polenta may explain incidents in medieval Italy. *ASM News* **60: 187-91**

d'Enfert, C., Reyss, I., Wandersmann, C and Pugsley, A.P. (1989)Protein sectretion by Gram-negative bacteria: characterisation of two membrane proteins required for pullulanase secretion by *E. coli* K12. *J Biol Chem* **264(29):** 17462-17468

Dale, J.W. (1994) *Molecular Genetics of Bacteria* 2nd Edtn. Chapter 4: Regulation of Gene Expression. pp 75-105. John Wiley and Sons, Chichester.

Daschner, F.D. (1980) Epidemiology of *Serratia marcescens*. pp187-194. *In* Von Graevenitz, A. and Rubin, S.J. (eds) *The genus* Serratia. CRC Press Inc, Boca Raton, Florida

Dauenhauer S.A . Hull, R,A & Williams, R,P (1984) Cloning and expression in *Escherichia .coli* of *Serratia marcescens* genes encoding prodigiosin biosynthesis. *J Bacteriol* **158(3):**1128-1132

Davies (1990) What are antibiotics? Archaic functions for modern activities. *Mol Microbiol* **4:** 1227-1232

de Vries, G.E., Raymond, CK., and Ludwig, R.A. (1984) Extension of bacteriophage λ host range: Selection, cloning and characterisation of a constitutive λ receptor gene. *Proc Nat Acad Sci* USA. **81:** 6080-6084

delCardayre S.B., Stock, K.P., Newton, G.L., Fahey, R.C. and Davies, J.E. (1998) Coenzyme A disulfide reductase: purification and characterization of the enzyme component of the primary thiol/disulfide redox system in *Staphylococcus aureus*. *J Biol Chem* (In press)

Demain, A.L. (1995) Why do microorganisms produce antimicrobials? pp205-228. *In* Hunter, P.A., Darby, G.K., and Russell, N.J (eds). *Fifty years of antimicrobials: past perspectives and future trends.* Society for General Microbiology. Symposium 53. Society for General Microbiology, Cambridge.

Demain, A.L. and Piret, J.M. (1979) Relationship between antibiotic biosynthesis and sporulation. pp183-188. *In* Lucker, M and Schreiber, K. (eds). *Regulation of secondary product and plant hormone metabolism.* Pergamon Press, New York

Diaz-Guerra, L., Moren, F. and San Millan, J.L. (1989) *appR* gene product activates transcription of microcin C7 plasmid genes. *J Bacteriol* **171(5):** 290-298

Ding M., and Williams, R.P. (1983) Biosynthesis of prodigiosin in white strains of *Serratia marcescens* isolated from patients. *J Clin Microbiol* **17(3):** 476-480

Ding Q, Kusano S, Villarejo M, Ishihama A (1995) Promoter selectivity control of *Escherichia coli* RNA polymerase by ionic strength: differential recognition of osmoregulated promoters by E sigmaD and E sigmaS holoenzymes. *Mol Microbiol* **16(4):** 649-656

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L. (1991) Molecular genetic organisation of genes required for complex polyketide synthesis. *Science* **252:** 675-679

Dorman, C.J. (1993) *Genetics of bacterial virulence*. Chapter 6: Transcriptional regulation. pp 139-169. Blackwell Scientific Publications, Oxford.

Edelman, A., Bowler, L., Broome-Smith, J.K. and Spratt, B.G. (1987) Use of a beta-lactamase fusion vector to investigate the organization of penicillin-binding-protein 1B in the cytoplasmic membrane of *Escherichia coli. Mol Microbiol* **1:** 101-106

Egger, L.A. and Inouye, M. (1997) Signal transduction via the histidyl-aspartyl phosphorelay. *Genes Dev* **2(3):** 167-184

Erlich, H.A. (ed.) (1989) *PCR Technology*. New York: Stockton Press

Feistner, G.(1990) Pigments.pp233-244. In Klement, Z., Rudolph, K., and Sands, D.C.(eds). *Methods in Phytobacteriology*. Akademiai Kiado, Budapest.

Feitelson, J.S. and Hopwood, D.A. (1983) Cloning of a *Streptomyces* gene for an *O*-methyl transferase involved in antibiotic biosynthesis. *Mol Gen Genet* **190:** 394-398

Feitelson, J.S., Malpartida, F., and Hopwood, D.A. (1985) Genetic and biochemical characterisation of the *red* gene cluster of *Streptomyces coelicolor* A3(2). *J Gen Microbiol.* **131:** 2431-2441

Fernández-Moreno, M.A., Caballero, J.L., Hopwood, D.A., and Malpartida, F. (1991) The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* transfer RNA gene of *Streptomyces. Cell.* **66:** 769-780

Fishman, S.E., Cox, K., Larson, J.L., Reynolds, P.A., Seno, E.T., Yeh, W.K., Van Frank., R. and Heshberger, C.L. (1987) Cloning genes for biosynthesis of a macrolide antibiotic. *Proc Natl Acad Sci* USA **23:** 8248-8252

Flavier, A.B., Schell, M.A. and Denny, T.P. (1998) An RpoS (sigma-S) homologue regulates acylhomoserine lactone-dependent autoinduction in *Ralstonia solancearum. Mol Microbiol* **28(3):** 475-486

Flaxman, C.S. (1995) A study of the regulation of undecylprodigiosin biosynthesis in *Streptomyces coelicolor* A3(2). PhD Thesis, University of Warwick

Forbes, K.J. and Perombelon, M.C.M. (1985) Chromosomal mapping of *Erwinia carotovora* subspecies *carotovora* with the IncP plasmid R68::Mu. *J Bacteriol* **164:** 1110-1116

Fuqua, W.C., Winans, S.C and Greenberg, E.P. (1994) Quorum sensing in bacteria: the LuxI-LuxR family of cell density-responsive transcriptional activators. *J Bacteriol* **176(2):** 269-275

Gardan, R., Rapoport, G. and Débarbouillé (1995) Expression of the *rocDEF* operon involved in arginine catabolism in *Bacillus subtilis. J Mol Biol* **249:** 843-856

Gaughran, E.R.L. (1969) From superstition to science: the history of a bacterium. *Trans N Y Acad Sci.***31:** 3-24

Gerber, N.N. (1975) Prodigiosin-like pigments. CRC *Crit Revs Microbiol* **462:** 469-485

Gerber, N.N. and Lechevalier, M.P. (1976) Prodiginine (prodigiosin-like) pigments from *Streptomyces* and other aerobic Actinomycetes. *Can J Microbiol.* **22:** 658-667.

Gish, W. and States, D.J. (1993) Identification of protein coding regions by database similarity search. *Nature Genet* **3:** 266-272

Gloecker, R., Ohsawa, I., Speck, D., Ledoux, C., Bernard, S., Zinsius, M., Villeval, D., Kisou, T., Komogawa, K. and Lemoine, Y. (1990) Cloning and characterization of the *Bacillus sphaericaus* genes controlling the bioconversion of pimelate into dethiobiotin. *Gene* **87(1):**63-70

Gold, L. and Stormo, G. (1987) Translation initiation. In *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. pp1302-1307. Niedhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaecter M. and Umbarger, H.E. (Eds). Am Soc for Microbiol, Washington DC

Gray, K.M. (1997) Intercellular communication and group behaviour in bacteria. Trends Microbiol 5(5): 184-188

Grimont, P.A.D., and Grimont, F. (1978) The Genus *Serratia. Annu Rev Microbiol.***32:** 221-248

Grimont, P.A.D., and Grimont, F. (1984) Genus VIII. *Serratia, In* Krieg, N.R., and Holt, J.G. (eds), Bergy's manual of systematic bacteriology, vol1. Williams and Wilkins Co., Baltimore.

Grimont, P.A.D., Grimont, F., Dulong de Rosnay, H.L.C. and Sneath, P.H.A. (1977) Taxonomy of the *genus Serratia. J Gen Microbiol* **98:** 39-48

Guthrie, E.P. and Chater, K.F. (1990) The level of a transcript required for production of a *Streptomyces coelicolor* Antibiotic is Conditionally Dependent on a tRNA gene. *J Bacteriol* **172(11):** 6189-6193

Guthrie, E.P., Flaxman, C.S., White, J., Hodgson, D.A., Bibb, M.J. and Chater, K.F. (1998) A response-regulator-like activator of antibiotic biosynthesis from *Streptomyces coelicolor* A3(2) with an amino-terminal domain that lacks a phosphorylation pocket. *Microbiol* **144(3):**727-738

Hara, O., and Beppu, T. (1982) Mutants blocked in streptomycin production in *Streptomyces griseus* - the role of A-factor. *J Antibiot.* **35:** 349-358

Harkki, A. and Palva, E.T. (1985) A *lamB* expression plasmid for extending the host range of λ to other enerobacteria. *FEMS Mic Lett* **27:** 183-187

Harris, S.J., Shih, Y.L., Bentley, S.D. and Salmond, G.P.C.S. (1998) The *hexA* gene of *Erwinia carotovora* encodes a LysR homologue and regulates motility and the expression of multiple virulence determinants. *Mol Microbiol* **28(4):**705-717

Harshey, R.M. (1994) Bees aren't the only ones: swarming in Gram negative bacteria. *Mol Microbiol* **13(3):** 389-394

Hearn, W.R., Elson, M.K., Williams, R.H. and Medina-Castro, J. (1970) Prodigiosene [5-(pyrryl-2,2'-dipyrrylmethene] and some substituted prodigiosenes. *J Org Chem* **35:** 142-144

Heinemann, B., Howard, A.and Palocz, H.J. (1970) Influence of dissolved oxygen levels on production of L-asparaginase and prodigiosin in *Serratia marcescens. Appl Micobiol* **19(5):** 800-804

Hejazi, A. and Falkiner, F.R. (1997) *Serratia marcescens. J Med Microbiol* **46(11):** 903-912

Helmann, J.D. and Chamberlin, M.J. (1988) Structure and function of bacterial sigma factors. *Annu Rev Microbiol* **57:** 839-872

Hines, D.A., Saurugger, P.N., Ihler, G.M., and Benedik, M.J. (1988) Genetic analysis of extracellular proteins of *Serratia marcescens. J Bacteriol* **170(9):** 4141-4146

Hood, D.W., Heidstra, R., Swoboda, U.K., and Hodgson, D.A. (1992) Molecular genetic analysis of proline and tryptophan biosynthesis in *Streptomyces coelicolor* A3(2): interactions between primary and secondary metabolism. *Gene* **115:** 5-12

Hopwood, D.A. and Sherman, D.H (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* **24:** 37-66

Hopwood, D.A., Chater, K.F., and Bibb, M.J. (1995) Genetics of Antibiotic Production in *Streptomyces coelicolor* A3(2), a model streptomycete. chapter 2. In Vining, L.C. and Stuttard, C. (eds) *Genetics and Biochemistry of Antibiotic Production.* Butterworth-Heinemann, USA.

Horinouchi, S., and Beppu, T. (1994) A-factor as a microbial hormone that controls cellular differentiation and secondary metabolism in *Streptomyces griseus. Mol Microbiol.* **12(6):** 859-864

Hsing, W. and Silhavy, T.J. (1997) Function of conserved histidine-243 in phosphatase activity of EnvZ, the sensor for porin osmoregulation in *Escherichia coli. J Bacteriol* **179(11):** 3729-3735

Hubert, E.G., Potter, G.M., Kalmanson, G.M. and Guze, L.B. (1969) Pigment in L-forms of *Serratia marcescens. J Gen Microbiol* **55:** 165-167

Hussain N.H., Goodson, M. and Rowbury, R.J. (1998) Recent advances in biology: intercellular communication and quorum sensing in micro-organisms. *Sci Prog 81(1):* 69-80

Ikatura, K., Rossi, J.J. and Wallace, R.B. (1984) Synthesis and use of synthetic oligonucleotides. *Annu Rev Biochem* **53:** 323-356.

Ishihama, A. (1993) Protein-protein communication within the transcription apparatus. *J Bacteriol* **175(9):** 2483-2489

Ishimoto KS, Lory S. (1989) Formation of pilin in *Pseudomonas aeruginosa* requires the alternative sigma factor (RpoN) of RNA polymerase. *Proc Natl Acad Sci* U S A **86(6):**1954-1957

Jung, J.U., Gutuerrez, C., Martin, F., Ardourel, M. and Villarejo, M. (1990) Transcription of *osmB*, a gene encoding *Escherichia coli* lipoprotein is

regulated by dual signals: osmotic stress and stationary phase. *J Biol Chem* **265(18):** 10574-10581

Kamarudin, M.I., Fox, L.K., Gaskins, C.T. and Gay, J.M. (1991) Environmental reservoirs for *Serratia marcescens*: intramammary infections in dairy cows. *J Amer Vet Assoc* **208(4):**553-558

Katoaka, T., Magae, J., Narauichi, H., Yamasaki, M. and Nagai, K. (1992) Enhancement by Concavalin A of the suppressive effects of prodigiosin-25C on proliferation of murine splenocytes. *J Antibiot* **45(8):** 1303-1312

Kawula, T.H. and Lelivelt, M.J. (1994) Mutations in a gene encoding a new Hsp70 supress rapid DNA inversion and *bgl* activation, but not *proU* derepression in *hns-1* mutant *Escherichia coli. J Bacteriol* **176(1):** 610-619

Kell, D.B., Kaprelyants, A.S., and Grafen, A. (1995) Pheromones, social behaviour and the functions of secondary metabolism in bacteria. *TREE.* **10:** 126-129

Killham, K. and Firestone,M.K. (1984a) Salt stress control of intracellular solutes in *Streptomycetes* indigenous to saline soils. *Appl Env Microbiol* **47:** 301-306

Killham, K. and Firestone,M.K. (1984b) Proline transport increases growth effects in salt-stressed *Streptomyces griseus. Appl Env Microbiol* **48:** 239-241

Kleerebezum, M., Quadri, L.E., Kupers, O.P. and de Vos, W.M. (1997) Quorum sensing by peptide pheromones and two-component signal transduction systems in Gram-positive bacteria. *Mol Microbiol* **24(5):** 895-904

Klenk, H.P., Clayton, R.A., Tomb, J., Nelson, K.E., White, O., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kervalage, A.R., Graham, D.E., Kyrpides, N.C., Fleischman, R.D., Quackenbush, J., Lee, N.H., Sitton, G.G., Gill, S., Kirkness E.F., Dougherty, B.A., McKenny, K., Adams, M.D., Loftus, B., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes S.M., Sadow, P.W., d'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen G.J., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter J.C. (1997) The complete genome sequence of the hyperthermophilic sulphate-reducing archaeon *Archaeoglobus fulgidus. Nature* **390(58):** 340-370

Kobayashi N. and Ichikawa, Y. (1989) A protein associated with prodigiosin formation in *Serratia marcescens. Microbiol Immunol* **33(4):** 257-263

Kondo, H., Nakagawa, A., Nishihira, J., Nishimura, Y., Mizuno, T and Tanaka, I. (1997) *Escherichia coli* positive regulator OmpR has a large loop structure at the putative RNA ploymerase interaction site. *Nat Struct Biol* **4(1):** 28-31

Korneluk, R.G., Quan, F. and Gravel, R.A. (1985) Rapid and reliable dideoxy sequencing of double-stranded DNA. *Gene* **40(2-3):** 317-323

Kraetzschmar, J., Krause, M. and Marahiel, M.A. (1989) Gramicidin S biosynthesis operon containing the structural genes *gsrA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases. *J Bacteriol* **171(10):** 5422-5429

Kyte, J. And Doolittle (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157:** 105-132.

Lange, R. and Hengge-Aronis, R. (1991) Growth-phase regulated expression of *bolA* and morphology of stationary-phase *Escherichia coli* cells are controlled by the novel sigma factor sigma S. *J Bacteriol* **173(14):** 4474-4481

Lange, R., Barth, M. and Hengge-Aronis, R. (1991) Complex transcriptional control of the sigma-S dependent stationary-phase-induced ans osmotically regulated *osmY* (*csi-5*) gene suggests novel roles for Lrp, cyclic AMP (cAMP) receptor protein-cAMP complex, and host integration factor in the stationary-phase response of *Escherichia coli.* *J Bacteriol* **175(24):** 7910-7917

Lauferska, U., Vina, M., Loren, J.G. and Guineau, J. (1983) Enhancement by polymyxin B of proline-induced prodigiosin biosynthesis in non-proliferating cells of *Serratia marcescens* *Microbiologica* **6(2):** 155-162

Lawanson, A.O and Sholeye, F.O. (1976) Inhibition of prodigiosin formation by adenosine triphosphate. *Experientia* **32:** 439-

Lawlor, E.J., Baylis, H.A., and Chater, K.F. (1987) Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). *Genes Dev* **1:** 1305-1310

Lee, M.H., Yamashita, M., Tsuji, R.F., Yamasaki, M., Katoaka, T, Magae, J., and Nagai, K. (1997) Suppression of T-cell stumilating function of allogeneic antigen presenting cells by prodigiosin 25-C. *J Antibiot* (Tokyo) **51(1):** 92-94

Lelivelt, M.J. and Kawula, T.H. (1995) Hsc66, an HSP70 homolog in *Escherichia coli*, is induced by cold shock but not by heat shock. *J Bateriol* **177(17):** 4900-4907

Lilljeström, P., Laamenan, I. and Palva E.T. (1988) Structure and expression of the *ompB* operon, the regulatory locus for the outer membrane porin regulon in *Salmonella typhimurium* LT-2. *J Mol Biol.* **201(4):** 663-673.

Lim, D,V., Qadri, S.M.H. and Williams, R.P. (1976) Incorporation of proline into prodigiosin by a Put mutant of *Serratia marcescens.* *Appl Environ Microbiol* **31:** 738-743

Lugtenberg, B., Peters, R., Bernheimer, H. and Berensden, W. (1976) Influence of culture conditions and mutations on the composition of outer membrane proteins of *Escherichia coli.* *Mol Gen Genet* **147(3):** 251-262

Magae, J., Miller, M.W., Nagai, K. and Shearer, G.M. (1996) Effects of metacycloprodigiosin, an inhibitor of killer T cells on murine skin and heart transplants. *J Antibiot* (Tokyo) **49(1):** 86-90

Maloy, S.R., Cronan Jr., J.E. and Freifelder, D. (1994a) Chapter 6: Gene Expression. In *Microbial Genetics*. 2nd Edtn. pp101-120. Jones and Bartlett, Boston, USA.

Maloy, S.R., Cronan Jr., J.E. and Freifelder, D. (1994a) Chapter 10: Mutagenesis, Mutations and Mutants. In *Microbial Genetics*. 2nd Edtn.. pp179-209. Jones and Bartlett, Boston, USA.

Malpartida, F., Niemi, J., Navarrete, R., and Hopwood, D.A. (1990) Cloning and expression in a heterologous host of the complete set of genes for the biosynthesis of the *Streptomyces coelicolor* antibiotic undecylprodigiosin. *Gene.* **93:** 91-99

Manoil, C. and Beckwith, J. (1985) Tn*phoA*: a transposon probe for protein export signals. *Proc Natl Acad Sci* USA **82:** 8129-8133

Manoil, C., Mekalanos, J.J. and Beckwith, J. (1990) Alkaline phosphatase fusions: sensors of subcellular location. *J Bacteriol* **172:** 515-518

Martín, J.F. and Liras, P. (1989) Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annu Rev Microbiol* **43:** 173-206

Martín, J.F., (1992) Clusters of genes for the biosynthesis of antibiotics: regulatory genes and overproduction of pharmaceuticals. *J Indust Microbiol* **9:** 73-90

Martinez-Hackert, E. and Stock., A.M. (1997) Structural relationships in the OmpR family of winged-helix transcription factors. *J Mol Biol* **13(3):** 301-312

Masters, M. (1985) Generalised Transduction. In Scaife, J., Leach, D. and Galizzi, A. (eds). *Genetics of Bacteria.* pp197-215. Academic Press, London

Matsuyama, T., Bhasin, A., and Harshey, R.M. (1995) Mutational Analysis of Flagellum-Independent Surface Spreading of *Serratia marcescens* 274 on a Low-Agar medium. *J Bacteriol.* **177(4):** 987-991

McGowan, S., Sebaihia, M., Jones, S., Yu, B., Bainton, N., Chan, P.F., Bycoft, B., Stewart, G.S.A.B., Williams, P. and Salmond, G.P.C. (1995) Carbapenem antibiotic production in *Erwinia carotovora* is regulated by CarR, a homologue of the LuxR transcriptional activator. *Microbiol* **141:** 541-550

Merrick, M.J (1993) In a class of its own - the RNA polymerase sigma factor sigma54 (sigma N). *Mol Microbiol* **10(5):** 903-909

Merrick, M.J. (1976) A morphological and genetic mapping study of bald colony mutants of *Streptomyces coelicolor. J Gen Microbiol.* **96:** 299-315

Messenger, A.J.M. and Turner, J.M. (1981) Effect of secondary metabolite production on the growth rate and variability of a pseudomonad. Society for *General Microbiology Quarterly* **8:** 263-264

Messing J. and Vieira, J. (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19:** 269-276

Miller, J.F., Calos, M.P., Galas, D., Hofer, M., Buchel, D.E. and Miller-Hill, B (1980) Genetic analysis of transposition in the lac region of *E. coli. J Mol Biol* **144:** 1-18

Miller, J.H. (1972) Experiments in molecular genetics. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Mitchell, R.E. (1991) Implications of toxins in the ecology and evolution of plant pathogenic microorganisms: bacteria. *Experientia* **47:** 791-803

Montilla, R., Williams, R.P., Lorén, J.G. and Viñas, M. (1991) Lipopolysaccharide is the receptor for kappa phage in *Serratia marcescens. Antonie van Leeuwenhoek* **59:** 15-18

Mootz, H.D., Marahiel, M.A. (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol* **179(21):** 6843-6850

Moreno, F., San Millan, J.L., Hernandez-Choco, C. and Kolter, R. (1995) Chapter 11: Microcins. *In* Vining, L.C. (ed.). *Genetics and Biochemistry of Antibiotic Production.* Butterworth Heinemann, USA

Morikawa M., Ito, M. and Imanaka, T. (1992) Isolation of a new surfactin producer, *Bacillus pumilus* A-1 and cloning and nucleotide sequence of the regulator gene *psf-1*. *J Ferm Bioeng* **74(5):**255-261

Morrison, D.A. (1966) Prodigiosin synthesis in mutants of *Serratia marcescens. J Bacteriol.* **91:** 1599-1605

MukherjeeA., Cui, Y., Ma, W., Liu, Y., Ishihama, A., Eisenstark, A. and Chatterjee, A.K. (1998) RpoS (sigma-S) controls expression of *rsmA*, a global regulator of secondary metabolites, harpin and extracellular proteins in *Erwinia carotovora. J Bacteriol* **180(14):** 3629-3634

Mulvey M.R., Switala, J., Borys, A and Loewen, P.C. (1990) Regulation of transcription of *katE* and *katF* in *Escherichia coli. J Bacteriol* **172(12):** 6713-6720

Nakano, M.M, Corbell, N. and Besson, J. (1992) Isolation and characterization of *sfp*: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. *Mol Gen Genet* **232(2):**313-321

Narva, K.E., and Feitelson, J.S. (1990) Nucleotide Sequence and transcriptional Analysis of the *redD* locus of *Streptomyces coelicolor* A3(2). *J Bacteriol.* **172(1):** 326-333

Niersbach, M., Kreuzaler, F., Geerse, R.H., Poistma, P. and Hirsch, H.J. (1992) Cloning and nucleotide sequence of the *Escherichia coli* K-12 *ppsA* gene, encoding PEP synthase. *Mol Gen Genet* **232:** 332-336

Nixon, B.T., Ronson, C.W. and Ausbel, F.M. (1986) Two-component regulator systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. *Proc Natl Acad Sci* USA **82(20):** 7850-7854

Parachuri, D.K. and Harshey, R.M. (1987) Flagellar variation in *Serratia marcescens* is associated with colour variation. *J Bacteriol* **169:** 61-65

Parker, W.L., Rathum, M.L., Wells, J.S. Jnr., Trejo, W.H., Pricip, P.A. and Sykes, R.B. (1982) A simple carbapenem produced by species of *Serratia* and *Erwinia. J Antibiot* **35(6):** 653-660

Parment, P.A. (1997) The role of *Serratia marcescens* in soft contact lens associated ocular infections. A review. *Acta Ophthalmol Scand* **75(1):** 67-71

Patel, K.A., Mehta, A.M. and Dave, P.G. (1982) Prodigiosin, a component of kappa phage receptor complex in *Serratia marcescens*. *Microbios* **34:** 153-158

Pocalyko D.J., Carroll, L.J., Martin, B.M., Babitt P.C. and Dunaway-Mariano, D. (1990) Analysis of sequence homologies in plant and bacterial pyruvate phosphate dikinase, enzyme I of the bacterial phosphoenolpyruvate: sugar phosphotransferase system and other PEP-utilizing enzymes. Identification of potential catalytic and regulatory motifs. *Biochemistry* **29:** 10757-10765

Poole, L.B. and Clairborne, A. (1989) The non-flavin redox center of the Streptococcal NADH peroxidase. *J. Biol Chem* **264:** 12322-12329

Pugsley, A.P. (1993). The Complete General Secretory Pathway in Gram-Negative Bacteria. *Microbiol Reviews.* **57(1):** 50-108

Qadri, S.M.H. and Williams, R.P. (1973) Role of methionine in biosynthesis of prodigiosin by *Serratia marcescens*. *J Bacteriol* **116(3):** 1191-1198

Randall-Hazelbauer and Schwartz, M. (1973) Isolation of the bacteriophage Lambda from *E. coli*. *J Bacteriol* **116(3):** 1191-1198

Redenbach, M., Keiser, H.M., Denapaite, D., Eichner, A. Cullum, J., Kinashi, H. and Hopwood, D.A. (1996) A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol Microbiol* **21(7):** 77-96

Reeves, P.J., Whitcombe, D., Wharam, S., Gibson, M., Allison, G., Bunce, N., Barallon, R., Douglas, P., Mullholland, V., Stevens, S., Walker, D., and Salmond, G.P.C. (1993) Molecular cloning and characterization of 13 *out* genes from *Erwinia carotovora* subspecies *carotovora:* genes encoding members of a general secretion pathway (GSP) widespread in Gram-negative bacteria. *Mol Microbiol* **8(3):** 443-456.

Reizer, J., Hoischen, C., Reizer, A., Pham, T.N. and Saier Jr., M.H. (1993) Sequence analyses and evolutionary relationships among the energy-coupling proteins Enzyme I and HPr of the bacterial phosphoenol pyruvate-sugar phosphotransferase system. *Protein Sci* **2(4):** 506-521

Revill, W.P. and Leadlay, P.F. (1991) Cloning, characterization and high-level expression in *Escherichia coli* of the *Saccharopolyspora erythraea* encoding acyl carrier protein potentially involved in fatty acid biosynthesis. *J Bacteriol* **173:** 4379-4385

Rius, N., Solé M., Francia A. and Lorén, J.G.(1994) Buffering capacity of pigmented and non-pigmented strains of *Serratia marcescens*. *Appl Environ Microbiol* **60(6):** 2152-2154

Rizki, M.T.N (1954) Diffusion of chromogenic indicators of *Serratia marcescens*. *Proc Natl Acad Sci* USA **40:** 1057-

Roberts, D.L., Bennett, D.W. and Forst, S.A. (1994) Identification of the phosphorylation site of the osmosensor EnvZ of *Escherichia coli*. *J Biol Chem* **269(12):** 8728-8733

Rosenberg, E. (1984) Isolation of pigmented and nonpigmented mutants of *Serratia marcescens* with reduced cell surface hydrophobicity. *J Bacteriol* **160(1):** 480-482

Rosenberg, E. and Varon, M. (1984) Antibiotics and Lytic Enzymes. *In* Rosenberg, E (ed.). *Myxobacteria Development and Cell Interactions.* pp104-125. Springer-Verlag, New York USA

Ross R.P. and Clairborne, A. (1991) Molecular cloning and analysis of the gene encoding the NADH oxidase from *Streptococcus faecalis* 10C1. Comparison with NADH peroxidase and the flavoprotein disulfide reductases *J Mol Biol* **221:** 658-671

Rudd, B.A.M., Hopwood, D.A. (1980) A pigmented mycelial antibiotic in *Streptomyces coelicolor* A3(2).: control by a chromosomal gene cluster. *J Gen Microbiol.* **119:** 333-340

Salmond G.P.C., Bycroft, B.W., Stewart G.S.A.B. and Williams, P. (1995) The bacterial 'enigma': cracking the code of cell-cell communication. *Mol Microbiol* **16:** 615-624.

Salmond, G.P.C., Hinton, J.C.D., Gill, D.R., and Perombelon, C.M. (1986) Transposon mutagenesis of *Erwinia* using phage λ vectors. *Mol Gen Genet.* **203:** 524-528.

Sambrook J., Fritsch, E.F., and Maniatis, T,. (1989) *Molecular Cloning a Laboratory Manual*. Cold Spring Harbour New York.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* USA **74(12):** 5463-5467

Sasakawa, C., Lowe, J.B., McDivitt, L and Berg, D.E. (1982) Control of Tn*5* transposition in *E. coli*. *Proc Natl Acad Sci* USA **79:** 7450-7474

Schell, M. (1993) Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol* **47:**597-626

Schwartz, M. and LeMinor, L. (1975) Occurrence of bacteriophage Lambda receptor in some Enterobacteriaceae. *J Virol* **15:** 679-685

Scott R., Qadri, S.M.H and Williams, R.P. (1976) Role of L-proline in the biosynthesis of prodigiosin. *Appl Environ Microbiol* **32:** 561-567

Seaton, B.L. and Vickery, L.E. (1994) A gene encoding a *DnaK/hsp70* homolog in *Escherichia coli*. *Proc Natl Acad Sci* USA **91:** 2066-2070

Selvaraij, G., Fong, Y.C. and Iyer, V.N. (1984) A PortableDNA sequence Carrying the Cohesive site (*cos*) of Bacteriophage l and the *mob* (mobilisation) Region of the Broad-Host-Range Plasmid RK2: a Module for the construction of New Cosmids. *Gene.* **32:** 235-241.

Shapiro, J.A. (1988) Bacteria as Multicellular Organisms. *Sci American* June: 62-69

Shaw, K.J. and Berg, C.M. (1979) *E. coli* K12 auxotrophs induced by the transposoable element Tn*5*. *Genetics* **92:** 741-744

Sherman, D.H., Malpartida, F., Bibb, M.J., Keiser, H.M. and Hopwood, D.A. (1989) Structure and deduced function of the granatacin-producing polyketide synthase gene of *Streptomyces violaceoruber* Tu22. *EMBO J* **8:** 2717-2725

Shine, J., and Dalgarno, J. (1974) The 3'-Terminal sequence of *E.coli* 16s RibosomalRNA: complementary to Nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA.* **71:** 1342-1346.

Siddiqui, M.A.Q. and Peterson, G.E. (1965) Streptomycin and syntrophic prodigiosin synthesis in *Serratia marcescens. Antonie van Leeuwenhoek* **31:** 193-

Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984) Electrophoresis of proteins. pp209-212. In *Experiments in Gene fusions*, Cold Spring HarborPress, New York

Simberkoff, M.S. (1980) Experimental *Serratia marcescens* infection and defense mechanisms. pp 158-163. *In* Von Graevenitz, A., and Rubin S.J. (eds) *The Genus* Serratia. CRC Press Inc, Boca Raton, Florida, USA

Sitnikov, D.M., Schineller, J.B. and Baldwin, T.O. (1996) Control of cell division in *E. coli*: regulation of transcription of *ftsQA* involves both *rpoS* and Sdi-A mediated autoinduction. *Proc Natl Acad Sci* USA **93(1):** 336-341

Sivendra, R. and Lo, H.S. (1975) Identification of *Chromobacteruim violaceum:* pigmented and non-pigmented strains. *J Gen Microbiol* **90(1):** 21-31

Slauch J.M. and Silhavy T.J. (1991) Genetic fusions as experimental tools. *Meth Enzymol* **204:** 213-248

Slauch, J.M., Russo, F.D., and Silhavy, T.J. (1991) Suppressor Mutations in *rpoA* Suggests that OmpR controls transcription by Direct interaction with the a subunit of RNA polymerase. *J Bacteriol.* **173(23):** 7501-7510

Smith D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J.N. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* **179(22):** 7135-7155

Solé, M., Rius, N., Francia, A., and Loren, J.G. (1994) The effect of pH on prodigiosin production by non-proiferating cells of *Serratia marcescens. Lett Appl Microbiol.* **19:** 341-344

Stone, M.J., Williams, D.H. (1992) On the evolution of functional secondary metabolites (natural products). *Mol Microbiol.* **6(1):** 29-34

Straus, D., Walter, W. and Gross, C.A. (1990) DnaK, DnaJ and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of Sigma-32. *Genes Dev* **4(12):** 2202-2209

Stryer, L.(1988a) *Biochemistry* 3rd Edtn. Chapter 12: Introduction to Biological Membranes. pp280-312. W.H. Freeman and Co. New York

Stryer, L.(1988b) *Biochemistry* 3rd Edtn. Chapter 20: Fatty Acid Metabolism. pp280-312. W.H. Freeman and Co. New York

Stryer, L.(1988c) *Biochemistry* 3rd Edtn. Chapter 21: Amino Acid degradation. pp280-312. W.H. Freeman and Co. New York

Summers R.G., Wendt-Pienkowski E., Motamedi H. and Hutchinson C.R. (1992) Nucleotide sequence of the *tcmII-tcm-IV* region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* and evidence that the tcmN gene encodes a multi-functional cyclase-dehydratase-O-methyl transferase. *J Bacteriol* **174(6):** 1810-1820

Syzbalski, W. (1993) From the double-helix to novel approaches to the sequencing of large genomes. *Gene* **35:** 279-290

Tabor, S. and Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci* USA **82:** 1074-1078

Taddyon, M., and Broome-Smith J.K. (1992) Tn*blaM*: a transposon for directly tagging bacterial genes encoding cell envelope secreted and proteins. *Gene* **111:** 21-26

Tanaka, K., Takayanagi, Y., Fujita, N., Ishihama, A. and Takahashi, H. (1993) Heterogeneity of the principal sigma factor in *Escherichia coli*: the *rpoS* gene product, sigma-38, is a second principal sigma factor of RNA polymerase in stationary-phase *Escherichia coli*. *Proc Natl Acad Sci* USA **90(8):** 3511-3515

Tang, X., Nakata, Y., Li, H., Zhang, M., Gao, H., Fujita, N., Ishihama, A., Sakasume, O., Ohta, T and Yokoyama, K. (1994) The optimization of preparations of competent cells for transformation of *E. coli*. *Nuc Acids Res* **22(14):** 2857-2858

Taylor, P.A., Schnoes, H.K. and Durbin, R.D. 1972 Characterisation of chlorosis-inducing toxins from a plant pathogenic *Pseudomonas* species. *Biochim Biophys Acta* **286(1):** 107-117

Thimann, K.V. (1963). *The Life of Bacteria* p.17. 2nd Edition, Macmillan Press, New York

Thomson, N.R. (1996) Coordinate regulation of antibiotic and pigment by the *Serratia* Rap protein: Evidence for a novel family of regulatory proteins. PhD Thesis, University of Warwick

Thomson, N.R., Cox, A., Bycroft, B.W., Stewart G.S.A.B., Williams, P. and Salmon, G.P.C. (1997) The Rap and Hor proteins of *Erwinia*, *Serratia* and *Yersinia*: a novel subgroup in a growing superfamily of proteins regulating diverse physiological processes in bacterial pathogens. *Mol Microbiol* **26(3):** 531-544

Tiffany, K.A., Roberts, D.L., Wang, M., Paschke, R., Mohsen, A.W., Vockley, J. and Kim J.J. (1997) Structure of human isovaleryl Co-A dehydrogenase at 2.6Å resolution: structural basis for specificity. *Biochemistry* **36(28):** 8455-8464

Tinoco, I., Uhlenbeck, O and Levine, M. (1973) Estimation of secondary structure in nucleic acids. *Nature* **230:**362-367

Tognoni, A., Franchi,. E., Magistrelli, C., Colombo, E., Cosmina, P., Grandi, G. (1995) A putative new peptide synthase operon in Bacillus subtilis: partial characterization. *Microbiology* **141(3):** 645-648

Tokishita, S., Kojima, A. and Minzuno, T. (1992) Transmembrane signal transduction and osmoregulation in *Escherichia coli*: function of the transmembrane regions of the membrane-located protein kinase EnvZ. *J Biochem* **111 (6):** 707-713

Totten PA, Lara JC, Lory S (1990) The *rpoN* gene product of *Pseudomonas aeruginosa* is required for expression of diverse genes, including the flagellin gene. *J Bacteriol* **172(1):**389-396

transcriptional activators. J Bacteriol 176: 1234-1241

Trutko, S.M., and Akimenko, V.K. (1988) The role of prodigiosin biosynthesis in the regulation of oxidative metabolism of the producer *Serratia marcescens* *Mikrobiologiya* **58(5):** 723-729

Tsang, J.C., Tattrie, S. and Kallvy, D. (1971) Outer envelope glycoprotein from two strains of *Serratia marcescens.* *Appl Microbiol* **21(1):** 27-31

Tsao , S.W., Rudd, B.A.M., He, X.G., Chang, C.J. and Heinz, H.G. (1985) Identification of a red pigment from *Streptomyces coelicolor* A3(2) as a mixture of prodigiosin derivatives. *J Antibiot* **38(1):** 128-131

Tsuji, R.F., Magae, J., Yamashita, M., Nagai, K., and Yamasaki, M. (1992) Immunomodulating properties of prodigiosin 25-C. An antibiotic which preferentially suppresses induction of cytotoxic T cells. *J Antibiotics.* **45:** 1295-1302

Van der Mei, H.C., Cowan, M.M., Genet, M.J., Rouxhet, P.G. and Busscher, H.J. (1991) Structural and physiochemical properties of *Serratia marcescens* strains. *Can J Microbiol* **38:** 1033-1041

van Ogtrop, M.L., van Zoeren-Groebben, D., Verbakel-Salamons, E.M., van Boven, C.P. (1997) *Serratia marcescens* infections in neonatal departments: description of an outbreak and a review of the literature. *J Hosp Infect* **36(2):** 95-103

Venturi, V., Weisbeek, P. and Koster, M. (1995) Genetic regulation of siderophore mediated iron-acquisition in *Pseudomonas*: not only the Fur repressor. *Mol Microbiol* **15(6):** 1081-1093

Vicente, M., Kushner, S.R., Garrido, T., Aldea, M. (1991) The role of the 'gearbox' in the transcription of essential genes. *Mol Microbiol* **5(9):** 2085-2091

Vickery, L.E., Silberg, J.J. and Ta, D.T. (1997) Hsc66 and Hsc20, a new heat shock cognate molecular chaperone system from *Escherichia coli.* *Protein Sci* **6(5):** 1047-1056

Vining, L.C. (1990) Functions of Secondary Metabolites. *Annu Rev Microbiol.* **44:** 395-427

Vining, L.C. (1992) Secondary metabolism, inventive evolution and biochemical diversity - a review. *Gene* **5:** 135-140

Von Heijne, G. (1986) Signal sequence recognition. *Nucl Acids Res.* **14:**4683

Vujaklija, D., Horinuchi, S. and Beppu, T. (1993) Detection of an A-factor responsive protein that binds to the upstream activation sequence of *strR*, a regulatory gene for streptomycin biosynthesis in *Streptomyces griseus. J Bacteriol* **175:** 2652-2661

Wang, R.C., Seror, S.J., Blight, M., Pratt,, J.M., Broome-Simth, J.K. and Holland, I.B. (1991) Analysis of the membrane organization of an Escherichia protein translocator, HlyB, a member of a large family of prokaryote and eukaryote surface transport proteins. *J Mol Biol.* **217(3):** 441-454.

Wasserman, H.H., Sykes, R.J., Peverada, P., Shaw, C.K., Cushley, R.J., and Lipsky, S.R. (1973) Biosynthesis of prodigiosin. Incorporation patterns of ^{13}C labelled alanine, proline, glycine and serine elucidated by fourier transform nuclear magnetic resonance. *J Am Chem Soc.* **95:** 6874-6876

Wehrli, W., Knusel, F., Schmid, K. and Staehelin, M. (1986a) Interaction of rifampicin with bacterial RNA polymerase. *Proc Natl Acad Sci* USA **61:** 667-673

Wehrli, W., Knusel, F., Schmid, K. and Staehelin, M. (1986b) Action of rifampicins on RNA polymerase. *Biochim Biophy Acta* **157:** 215-217

White, J. and Bibb, M. (1997) *bldA* dependence of undecylprodigiosin in *Streptomyces coelicolor* A3(2) involves a pathway specific regulatory cascade

Williams R.P. & Quadri S.M. (1980) The pigment of *Serratia* p 31-75 in Von Graevenitz, A., and Rubin S.J. (eds) *The Genus* Serratia. CRC Press Inc, Boca Raton FL.

Williams, D.H.., Stone, M.J., Hauck, P.R., and Rahman, S.K. (1989) Why are secondary metabolites biosynthesised? *J Natl Prod.* **52:** 1189-1208

Williams, R.P. (1973) Biosynthesis of prodigiosin, a secondary metabolite of *Serratia marcescens. Appl Microbiol* **25(3):** 396-402

Williams, R.P. and Gott, C.L. (1964) Inhibition by streptomycin of the biosynthesis of prodigiosin. *Biochem Biophy Res Commun* **16**: 47-51

Williams, R.P., Goldschmidt, M.E. and Gott, C.L. (1965) Inhibition by temperature of the terminal step in biosynthesis of prodigiosin. *Bioch Biophys Res Comm* **19:** 177-182

Williams, R.P., Gott, C.L. and Qadri, S.M.H.(1971a) Induction of pigmentation in non-proliferating sells of *Serratia marcescens* by addition of single amino acids. *J Bacteriol* **106(3):** 444-448

Williams, R.P., Gott, C.L., Qadri, S.M.H., and Scott, R.H.(1971b) Influence of temperature of incubation and type of growth media on pigmentation in *Serratia marcescens. J Bacteriol.* **106(7):** 438-443

Williams, R.P., Scott, R.H., Lim, D.V. and Qadri, S.M.H. (1976) Macromolecular syntheses during biosynthesis of prodigiosin by *Serratia marcescens*. *Appl Env Microbiol* **31(1):** 70-77

Wilmes-Riesenberg, M.R., and Wanner, B.L. (1992). Tn*phoA* and Tn*phoA'* elements for making and switching fusions for study of transcription, translation, and cell surface localisation. *J Bacteriol.* **174(14):** 4558-4575

Wise, A., Brems, R., Ramakrishnan, V. and Villarejo, M. (1996) Sequences in the -35 region of *Escherichia coli rpoS*-dependent genes promote transcription by E-sigmaS. *J Bacteriol* **178(10):** 2785-2793

Witney, F.R., Failla, M.L., Weinberg, E.D. (1977) Phosphate inhibition of secondary metabolism in *Serratia marcescens*. *Appl Environ Microbiol* **33(5):** 1042-1046

Woods, D.R., Mosmann, R., Hanson, S., Hendry, D.A. (1971) Pigmentation and acraflavine resistance in *Serratia marcescens*. *J Bacteriol* **108(1):** 765-769

Wurtzel, E.T., Chou, M.Y., Inouye, M. (1982) Osmoregulation of gene expression: DNA sequence of the *ompR* gene if the *ompB* operon of *Escherichia coli* and characterization of its gene product. *Proc Natl Acad Sci* USA **257(22):** 13685-13691

Yanagida, N., Uozimi, T. and Beppu, T. (1986) Specific secretion of a *Serratia* protease through the outer membrane of *Escherichia coli*. *J Bacteriol* **166:** 937-944

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 cloining vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. *Gene* **33:** 103-119

Yim, H.H., Brems, R.I. and Villarejo, M. (1994) Molecular characterization of the promoter of *osmY*, a *rpoS*-dependent gene. *J Bacteriol* **176(1):** 100-107

Zahner, H. (1979) What are secondary metabolites? *Folia Microbiol* **24(5):** 435-443

Zhang, Y., Rowley, K.B. and Patil, S.S. (1993) Genetic organisation of a cluster of genes involved in the production of phaseolotoxin, a toxin produced by *Pseudomonas syrigae* pv. *phaseolica*. *J Bacteriol* **175(20):** 6451-6458